

THE SPECTRAL IMPACT OF THE HYPOPHARYNGEAL CAVITIES
ON THE SINGING VOICE

BERTRAND DELVAUX

Doctor of Philosophy

University of York
Electronics

January 2014

Bertrand Delvaux: *The spectral impact of the hypopharyngeal cavities on the singing voice*, © January 2014

The human voice is the organ of the soul.

— Henry Wadsworth Longfellow, American poet

Dedicated to my family and to Michael Spyres, my oversea brother
who inspired me to start this thesis.

ABSTRACT

This thesis investigates how the morphological features of the lower vocal tract impact the singing voice of professional singers. Within the limitations of the source-filter theory, the resonances of the vocal tract are investigated, first under the lens of theoretical predictions for one tube, then with numerical and experimental results of a series of vocal tract models, and eventually with MRI-based vocal tracts measured for a group of professional solo singers. The method used is a comparison between theoretical predictions, numerical simulations (Finite Volume Method and Finite Element Method) and experimental results (measurement of MRI-based 3D-printed vocal tracts). The results suggest a strong and explainable correlation between the morphological dimensions of the hypopharyngeal cavities and the vallecula and the voice classification of an individual singer.

CONTENTS

Abstract	iv
List of Figures	x
List of Tables	xxxi
Acknowledgments	xxxiv
Declaration	xxxv
i INTRODUCTION	1
1 HYPOTHESIS	2
1.1 Singularity of the singing voice	2
1.2 How the voice works	2
1.3 Hypothesis	4
1.4 Thesis Outline	7
1.5 Contributions	9
2 SINGING VOICE PRODUCTION	10
2.1 Anatomy of singing production	13
2.1.1 Phonation	14
2.1.2 Articulation	19
2.2 Source-Filter theory	19
2.2.1 Framework	20
2.2.2 The model	21
2.3 Vocal Tract	22
2.3.1 Origin the Vocal Tract resonances	23
2.4 Formants	27
2.4.1 Consequences for the human voice	28
2.4.2 The Singer's Formant Cluster	29
2.4.3 The Hypopharynx	35
2.5 Volumetric imaging	45
2.5.1 Vocal Tract Area	47

2.6	Studying the Vocal Tract resonances	48
2.6.1	Output sound after excitation at the glottis	49
2.6.2	Output sound after excitation at the lips	50
2.6.3	Numerical simulation of the transfer function of an image-based VT	51
2.6.4	Measurement of an image-based 3D-printed VT	51
2.6.5	Methods used in this thesis	51
2.7	Voice classification	52
ii	MATERIALS AND METHODS	56
3	NUMERICAL SIMULATIONS	63
3.1	Computational domain	63
3.2	Finite Volume Method	64
3.2.1	Wave equation	65
3.2.2	Impulse Response	65
3.2.3	Transfer function	67
3.3	Finite Element Method	68
3.3.1	Transfer function	68
3.4	Meshing	69
3.5	Block-Diagram	69
4	EXPERIMENTAL MEASUREMENTS	73
4.1	Method	73
4.2	Material	76
4.2.1	The driver	76
4.3	Exponential Sine Sweep	77
4.4	Linearisation of the impulse response	83
4.4.1	Note about the harmonic distortions	87
4.5	Fast Fourier Transform	89
4.6	Final transfer function	91
4.7	Note about ESS and ESS^{-1}	91
4.7.1	A smooth start	91
4.7.2	A smooth end	93
4.8	Benchmarking	97

4.8.1	Driver-independent	97
4.8.2	Distance dependence	98
4.8.3	Absorption coefficient dependence	98
5	FROM MRI TO 3D PRINTING ...	102
5.1	Magnetic Resonance Imaging	102
5.1.1	Acquisition and protocol	103
5.1.2	Segmentation	104
5.2	3D printing	111
5.2.1	Technical data	112
5.2.2	Pieces printed	113
iii	RESULTS	117
6	RESULTS	119
6.1	Theoretical predictions	120
6.1.1	VTM-1	120
6.1.2	VTM-2	123
6.2	VTM-2 : changing the dimensions of the epilaryngeal tube	130
6.2.1	Changing the length of the epilaryngeal tube	130
6.2.2	Changing the radius of the epilaryngeal tube	131
6.3	VTM-Ch&K	138
6.3.1	Changing the radius	138
6.3.2	Changing the length	140
6.4	VTM-MRI	151
6.4.1	Comparison across voice categories	151
6.4.2	Dimensions of the Vocal Tract	154
6.4.3	Effect of the epilarynx	159
6.4.4	Effect of the piriform fossae	178
6.4.5	Effect of the vallecula	188
6.4.6	Numerical versus experimental	190
6.5	A new metric to compare Singer's Formant Clusters	194
6.5.1	Test with prototypical cases	197
6.6	Data visualisation with the new metrics	200

6.6.1	Comparison with the dimensions	200
6.6.2	Experimental versus Numerical	201
6.6.3	Singers Classification	201
6.6.4	Same singer, different vowels	204
6.6.5	Registers	205
6.6.6	Singer hallmark	206
iv	CONCLUSION	208
7	CONCLUSION	209
7.1	Background	210
7.2	Brief summary of the method used	211
7.2.1	A novel transducer-independent experimental method	212
7.3	Summary of findings	212
7.3.1	Identification of the spectral impact of the epila-ryngeal tube on VTM-Ch&K	214
7.3.2	A new Metric for the SFC	214
7.3.3	A formula to predict the SFC frequency centre based on the hypopharynx dimensions	215
7.3.4	Findings	216
7.4	Discussion	217
7.5	Model for the hypopharyngeal cavities	218
7.6	Contributions	221
7.7	Shortcomings	221
7.8	Further development	222
v	APPENDIX	225
A	ACOUSTIC MODES FOR AN OPEN CYLINDER	226
B	FINITE VOLUME METHOD	228
B.1	Basic principles	229
B.2	Time discretisation	230
B.3	Space discretisation	231
B.4	Comparison with other methods	232
C	FINITE ELEMENT METHOD	234

c.1	Basic Principles	234
c.2	Garlekin Method	237
c.3	Poisson problem	238
c.3.1	Choice of the mesh, the nodes and the local fields	239
D	MRI	244
D.1	Basic principles	245
D.2	Acquisition and Protocol	249
	BIBLIOGRAPHY	251

LIST OF FIGURES

- Figure 1.1 Vocal Tract and hypopharynx cavities superimposed on MRI mid-sagittal slice 5
- Figure 1.2 Hypopharynx details. 6
- Figure 2.1 A sagittal view of the human speech/singing production, from [38] 14
- Figure 2.2 The entrance to the larynx, viewed from behind, Coronal section of larynx and upper part of trachea and Laryngoscopic view of interior of larynx. From Henry Gray (1825-1861). *Anatomy of the Human Body*. 1918. Figs 955, 954 and 956 16
- Figure 2.3 Vocal Folds vibratory cycle. The Vocal Folds are first held shut by muscular tension, but the increasing subglottal pressure eventually manages to open the sealed Vocal Folds. As air is passing through the glottis, the transglottal pressure decreases (Bernouilli's principle), causing a suction effect, causing the sealing of the Vocal Folds [102]. When the glottis is closed, the subglottal pressure is greater than the epiglottal pressure: if the vocal folds are adjusted for phonation, they cannot resist this gradient of pressure and open again [102]. 18
- Figure 2.4 Schematic of the Speech/Singing production system, from [38] 20

- Figure 2.5 A schematic of the source-filter model, after [28]. The glottal source spectrum is produced at the glottis, whether it is periodic (singing, upper left) or continuous (whispering, upper right). The continuous transfer function of the vocal tract, as well as the positive radiation impedance spectrum are superimposed to the glottal source spectrum to form the output sound spectrum, either in singing (bottom left) or in whispering (bottom right). 22
- Figure 2.6 Pressure of the 3 first modes of an open cylinder, i. e. for $\lambda_1 = 4L$ ($\Rightarrow L = \lambda_1/4$ in red), $\lambda_3 = 4L/3$ ($\Rightarrow L = 3\lambda_3/4$ in blue) and $\lambda_5 = 4L/5$ ($\Rightarrow L = 5\lambda_5/4$ in cyan). 26
- Figure 2.7 Figure taken from [119]. Part A shows successively a straightened Vocal Tract, emphasising the location of the different constrictions from the glottis to the lips, The conversion between a DC flow coming from the trachea and the lungs to an AC flow in the Vocal Tract, with successive reflections at the mouth and at the glottis, and then the first resonances of an open tube (closed at the glottis and opened at the lips). Part B gives the impedance and transfer function of a cylindrical vocal tract of length=170 mm and radius=15 mm with and without an appended narrow constricted glottis. 27
- Figure 2.8 Long time average spectrum (LTAS) of a tenor. A singing formant is apparent in the range of 2500-3500 Hz (from [99]). 29

- Figure 2.9 Idealised vocal tract modification to generate a singing formant. (a) Uniform tube (dashed) and the uniform tube with a narrowed epilarynx (solid). (b) Frequency response functions of both the vocal tract shapes in (a) using the same respective line styles; the gray line represents the frequency response of the epilaryngeal tube in isolation. Figure taken from [99]. 30
- Figure 2.10 The fundamental frequency f_0 versus the first resonance R1 for sopranos. When f_0 is of equal or greater value than R1, sopranos tune R1 to a value slightly above f_0 . Taken from [53] 34
- Figure 2.11 The epilarynx as a tube-in-a-tube. The tube-shaped epilarynx is found at the bottom of the pharynx tube; together these structures define the lower vocal tract. The action of these tubes are pharyngeal constriction (1) and epilaryngeal constriction (2). From [66]. 40
- Figure 2.12 Anatomical sketch of the larynx. Sagittal section (a); posterior view (b). A = arytenoid cartilage; C = cricoid cartilage; T = thyroid cartilage; VF = vocal fold. Illustrations of laryngeal anatomy important to the epilaryngeal tube. From [66], after [122]. 41
- Figure 2.13 Functional planes of the larynx. Arrows indicate the axes of activity in each plane, the opposed vertical arrows between the ventricular and glottis plane indicate that these planes make contact. The dashed lines inscribed on the ellipses indicate lines of stricture. Dimension: x-axis is lateromedial; y-axis is inferiosuperior; z-axis is posteroanterior. From [66]. 41

- Figure 2.14 Relationship between the SFC and the dimensions of the VT, from [25] 53
- Figure 2.15 Vocal Tract Models used in this thesis. By order of increasing complexity/realism, VTM-1, VTM-2, VTM-Ch&K and VTM-MRI. 58
- Figure 2.16 Diagram of the different VTMs used in the thesis. By order of complexity/realism, VTM-1 (1 tube), VTM-2 (2 tubes), VTM-Ch&K (based on Chiba & Kajiyama's [14]) and VTM-MRI, actual VT shape acquired by MRI of six professional singers. Note that Bartholomew's VT has been processed for 5 registers and Barnaby's VT for 5 vowels. 61
- Figure 3.1 Computational domain of VTM-1: separation of the boundary between two patches: the VT walls and the radiating cavity. 64
- Figure 3.2 Computational domain for VTM-MRI-Barnaby-/stern/, with the different boundary conditions (Vocal Tract walls and Radiating cavity). The virtual probe microphone is located 3 cm far from the centre of the cross-sectional area of the lips opening and the point source (in yellow) is located at the glottis. Note that in the case of the experimental measurement of the 3D printed VTM-MRI, the (transducer) source and the microphone are interchanged. 67

Figure 3.3 Block-diagram of the processes at work for the numerical simulations and 3D printing of VTM-1, VTM-2, VTM-Ch&K and VTM-MRI. The processes are divided into three categories: the pre-processing, the simulation and the 3D printing. For VTM-1, VTM-2 and VTM-Ch&K (blue arrows): a script in Python is written in Salome to create the different geometries and subsequent meshes, assigning boundary conditions to the patches which are then used both for FVM (OpenFOAM) or FEM (Actran), which implement the equation to be solved as well as the boundary and initial conditions on the computational domain. VTM-MRI follows the red arrows: ITK-Snap is first used to segment the MRI data from the singers, exported into Blender which inverts the face orientation, then to NetGen which automatically meshes the volume with tetrahedrons, then to Blender, which manually selects the faces to split the domain into patches which will be assigned different boundary conditions in Salome. 72

- Figure 4.1 An Exponential Sine Sweep (ESS) is given as an input signal to the driver (1). The output recorded via a probe microphone is convolved with the inverse filter (ESS^{-1}) (2). It results in a temporal separation of the Linear Impulse Response (LIR) and the harmonic distortions (3). An FFT is performed on the LIR to give the linear transfer function of the system (4). Processes 1 to 4 are repeated twice: once with the Vocal Tract, and once without. Both spectra are then subtracted (logarithmic vertical scale) to give the transfer function of the standalone Vocal Tract. 75
- Figure 4.2 6-sided anechoic chamber. 76
- Figure 4.3 Frequency response of the driver VISATON SC 8 N 8 Ω 76
- Figure 4.4 3D-printed cabinet, designed to enclose the VISATON driver 77
- Figure 4.5 Spectrogram of the system response to an exponential sine Sweep in the case of a weakly nonlinear system, displaying the linear response of the system (main line on the bottom right) and its harmonic distortions (parallel lines to the main line). The amplitude is represented by a colour scale. 81
- Figure 4.6 The spectrum of the exponential sine sweep displays an energy drop of -3dB/octave 84
- Figure 4.7 The spectrum of the time-reversed exponential sine sweep displays an energy drop of -3dB/octave 84
- Figure 4.8 The spectrum of the amplitude-modulated exponential sine sweep displays an energy rise of +3dB/octave 86

- Figure 4.9 Convolution of the system response with the inverse filter signal. The plots are spectrograms: the horizontal axis is the time and the vertical axis is the frequency. On the upper left part, the system response is plotted (the thick line is the linear response of the system and the parallel lines are the harmonic distortions). On the upper right part is the inverse filter of the input signal. Both are convolved in the time domain (horizontal axis). The convolution process is shown at different times (t_1, t_2, t_3, t_4) on the bottom part of the figure: the intersections of the system response with the inverse filter leads to the LIR split from its harmonic distortions (bottom right). 87
- Figure 4.10 Result after the convolution of the system response with the inverse filter: the harmonic distortions are packed before the linear response on the time axis, parallel to the latter. Scale is in dB to enhance visualisation. 88
- Figure 4.11 Spectrum of the LIR (in red) and its harmonic distortions (in blue). 89
- Figure 4.12 Algorithm used to obtain the transfer function out of an impulse response. 90
- Figure 4.13 The start of the sweep is not smooth. After applying the envelope described in (4.10) and (4.11), the signal goes smoothly from zero to the start of the sweep. 92
- Figure 4.14 The sweep does not stop smoothly. After applying the envelope described in (4.10) and (4.11), the signal goes smoothly from the end of the sweep to zero. 94

- Figure 4.15 An Exponential Sine Sweep (ESS) of the form (4.1) has a burst of energy across the whole spectrum both at its start and at its end (A₁). Once convolved with its inverse filter (A₂), it leads to an impulse response and its echoes in the frequency-time space (A₃). Providing a smooth start to the (ESS) (B₁), and convolving it with its inverse filter (B₂) removes the pre-ringing (B₃). Providing the (ESS) with both a smooth start and a smooth end (C₁), and convolving it with its inverse filter (C₂) removes both the pre- and the post-ringing (C₃). 96
- Figure 4.16 Driver-independent experimental method. Measured transfer functions of the resonance of VTM-1 with 2 different speakers: a 952.210UK driver unit (in blue) and a VISATON SC 8 N 8 Ω transducer (in red). 97
- Figure 4.17 Distance-dependence of the resonances of VTM-Ch&K-/a/. Simulated transfer functions of the resonances of VTM-Ch&K-/a/, when the distance between the lips opening and the virtual probe takes the values 1cm, 2cm, 3cm, 5cm and 10 cm. 98
- Figure 4.18 Distance-dependence of the resonances of VTM-MRI-Barnaby-/hard/. Simulated transfer functions of the resonances of VTM-MRI-Barnaby-/hard/, when the distance between the lips opening and the virtual probe takes the values 1cm, 2cm, 3cm, 5cm and 10 cm. 99
- Figure 4.19 Absorption-dependence of the resonances of VTM-Ch&K-/a/. The simulated transfer function of VTM-Ch&K-/a/ is plotted for different values of absorption coefficient ($\alpha = 0.001, 0.01, 0.03, 0.1$). 101

- Figure 4.20 Absorption-dependance of the resonances of VTM-MRI-Barnaby-/hard/. The simulated transfer function of VTM-MRI-Barnaby-/hard/ is plotted for different values of absorption coefficient ($\alpha = 0.001, 0.01, 0.03, 0.1$). 101
- Figure 5.1 MRI mid-sagittal slice. 107
- Figure 5.2 MRI sagittal slice pre-segmentation and intensity regions. 108
- Figure 5.3 Bubbles expansion process: bubbles are placed in the vocal tract and expand until they reach a boundary defined by the intensity region. 109
- Figure 5.4 Manual editing after semi-automatic segmentation. 110
- Figure 5.5 Manual editing for the teeth 111
- Figure 5.6 3D printed pieces for VTM-1 (oropharynx only) and VTM-2 (oropharynx + epilarynx). 115
- Figure 5.7 3D printed VTM-MRI for the vowels as in /neap/ and /port/. 116
- Figure 6.1 Theoretical predictions, numerical simulations (FEM) and experimental measurements of the resonances of one tube. On the upper part of **A**, the theoretical predictions are given as the roots of the equation (6.2), plotted in red. On the lower part of **A**, these are linked to the numerical simulation (FEM) in black and experimental result in grey in by grey dotted lines. **B** shows VTM-1 and its respective dimensions. 122

Figure 6.2 FEM numerical simulations of the resonances of one tube and a twin tube, showing the effect of appending an extra tube and generating an extra resonance. On **A**, in red, the resonances of the oro-pharyngeal tube (VTM-1): theoretical predictions (red triangles) and numerical simulation (red plot). In blue, the resonance of the epilaryngeal tube (asymptote). In black, the twin-tube (VTM-2) resonances: theoretical predictions (black diamonds) and numerical simulations (greyscale plot). **B** shows VTM-2, with the dimensions of the epilaryngeal and oro-pharyngeal tubes respectively. 125

Figure 6.3 FVM numerical simulations of the resonances of one tube and a twin tube, showing the effect of appending an extra tube and generating an extra resonance. On **A**, in red, the resonances of the oro-pharyngeal tube (VTM-1): theoretical predictions (red triangles) and numerical simulation (red plot). In blue, the resonance of the epilaryngeal tube (asymptote). In black, the twin-tube (VTM-2) resonances: theoretical predictions (black diamonds) and numerical simulations (greyscale plot). **B** shows VTM-2, with the dimensions of the epilaryngeal and oro-pharyngeal tubes respectively. 126

- Figure 6.4 Experimental results of the resonances of one tube and a twin tube, showing the effect of appending an extra tube and generating an extra resonance. On **A**, in red, the resonances of the oro-pharyngeal tube (VTM-1): theoretical predictions (red triangles) and experimental results (red plot). In blue, the resonance of the epilaryngeal tube (asymptote). In black, the twin-tube (VTM-2) resonances: theoretical predictions (black diamonds) and experimental results (greyscale plot). **B** shows VTM-2, with the dimensions of the epilaryngeal and oro-pharyngeal tubes respectively. 127
- Figure 6.5 FEM simulated transfer functions when the epilaryngeal tube length of VTM-2 varies from 16 mm to 28 mm (lower part of the figure). The upper part of the figure represents the first resonance of the standalone epilaryngeal tube, which decreases in frequency as the epilaryngeal tube length varies from 16 mm to 28 mm. 134
- Figure 6.6 FEM simulated transfer functions when the epilaryngeal tube radius of VTM-2 varies from 9 mm to 2 mm. In blue, the evolution of the first resonance frequency of the standalone epilaryngeal tube when its radius changes from 9 mm to 2 mm. 135
- Figure 6.7 FVM simulated transfer functions when the epilaryngeal tube radius of VTM-2 varies from 9 mm to 2 mm. In blue, the evolution of the first resonance frequency of the standalone epilaryngeal tube when its radius changes from 9 mm to 2 mm. 135

- Figure 6.8 FVM simulated transfer functions when the epilararyngeal tube length of VTM-2 varies from 16 mm to 28 mm (lower part of the figure). The upper part of the figure represents the first resonance of the standalone epilararyngeal tube, which decreases in frequency as the epilararyngeal tube length varies from 16 mm to 28 mm. 136
- Figure 6.9 Variation of the SFC amplitude (from Fig 6.7) in relation with the epilararyngeal radius in VTM-2. 137
- Figure 6.10 FVM simulated transfer functions (A) when changing the epilararyngeal tube radius in Chiba & Kajiyama VTM-Ch&K-/i/ B from 10 mm to 4 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilararyngeal tube. The blue area visually aids to see the increasing radiated power of the SFC as the epilararyngeal tube radius decreases. 143
- Figure 6.11 FEM simulated transfer functions (A) when changing the epilararyngeal tube length and radius in Chiba & Kajiyama VTM-Ch&K-/a/ B from 16 mm to 28 mm and from 9 mm to 3 mm respectively. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilararyngeal tube. 144
- Figure 6.12 FVM simulated transfer functions (A) when changing the epilararyngeal tube length in Chiba & Kajiyama VTM-Ch&K-/a/ B from 16 mm to 18 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilararyngeal tube. 145

- Figure 6.13 FVM simulated transfer functions (**A**) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/i/ **B** from 14.5 mm to 26.5 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube. 146
- Figure 6.14 FVM simulated transfer functions (**A**) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/a/ **B** from 15 mm to 27 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube. 147
- Figure 6.15 FVM simulated transfer functions (**A**) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/e/ **B** from 14 mm to 26 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube. 148
- Figure 6.16 FVM simulated transfer functions (**A**) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/o/ **B** from 14 mm to 26 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube. 149
- Figure 6.17 FEM simulated transfer functions for VTM-Ch&K. From top to bottom, VTM-Ch&K-/i/, VTM-Ch&K-/u/, VTM-Ch&K-/e/, VTM-Ch&K-/o/, VTM-Ch&K-/a/. 150
- Figure 6.18 FEM transfer functions of different voice types phonating on /hard/. From top to bottom, Barnaby, Bartholomew, Timothy, Marisa, Maristela, Sophy. Details of the phonation and resonance data are given in Table 6.1. 153

- Figure 6.19 Variation of the SFC frequency centre predicted by 6.12 in relation with the Epilarynx Length across singers: **Soprano Sophy**, **Mezzo-Soprano 1 Maristela**, **Mezzo-Soprano 2 Marisa**, **Tenor Timothy**, **Bari-Tenor Bartholomew** and **Bass-Baritone Barnaby**. 157
- Figure 6.20 Epilaryngeal length (EL) and Pharyngeal width (PW) in the suggested formula 6.12 to predict the SFC centre. 158
- Figure 6.21 FEM simulated transfer function of VTM-MRI-Sophy-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 161
- Figure 6.22 FEM simulated transfer function of VTM-MRI-Marisa-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 161
- Figure 6.23 FEM simulated transfer function of VTM-MRI-Maristela-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 162
- Figure 6.24 FEM simulated transfer function of VTM-MRI-Timothy-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 162

- Figure 6.25 FEM simulated transfer function of VTM-MRI-Bartholomew-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 163
- Figure 6.26 FEM simulated transfer function of VTM-MRI-Barnaby-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 163
- Figure 6.27 FEM simulated transfer function of VTM-MRI-Barnaby-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 164
- Figure 6.28 FEM simulated transfer function of VTM-MRI-Barnaby-/port/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 164
- Figure 6.29 FEM simulated transfer function of VTM-MRI-Barnaby-/stern/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 165
- Figure 6.30 FEM simulated transfer function of VTM-MRI-Barnaby-/food/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 165

- Figure 6.31 FEM simulated transfer function of VTM-MRI-Barnaby-/neap/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx. 166
- Figure 6.32 FEM simulated transfer functions of VTM-MRI-Sophy-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation. 168
- Figure 6.33 FEM simulated transfer functions of VTM-MRI-Marisa-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation. 169
- Figure 6.34 FEM simulated transfer functions of VTM-MRI-Maristela-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation. 169
- Figure 6.35 FEM simulated transfer functions of VTM-MRI-Timothy-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation. 170

- Figure 6.36 FEM simulated transfer functions of VTM-MRI-Bartholomew-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation. 170
- Figure 6.37 FEM simulated transfer functions of VTM-MRI-Barnaby-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation. 171
- Figure 6.38 Five first resonance frequency values for VTM-MRI-Sophy-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.32 from the top. 175
- Figure 6.39 Five first resonance frequency values for VTM-MRI-Sophy-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.33 from the top. 175
- Figure 6.40 Five first resonance frequency values for VTM-MRI-Maristela-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.34 from the top. 176

- Figure 6.41 Five first resonance frequency values for VTM-MRI-Timothy-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.35 from the top. 176
- Figure 6.42 Five first resonance frequency values for VTM-MRI-Bartholomew-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.36 from the top. 177
- Figure 6.43 Five first resonance frequency values for VTM-MRI-Barnaby-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.37 from the top. 177
- Figure 6.44 Experimental results for VTM-MRI-Barnaby-/food/ with (grayscale) and without left (red) or right (blue) piriform fossa. 182
- Figure 6.45 Experimental results for VTM-MRI-Barnaby-/neap/ with (grayscale) and without left (red) or right (blue) piriform fossa. 182
- Figure 6.46 Numerical results for VTM-MRI-Sophy-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 183
- Figure 6.47 Numerical results for VTM-MRI-Marisa-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 183

- Figure 6.48 Numerical results for VTM-MRI-Maristela-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 184
- Figure 6.49 Numerical results for VTM-MRI-Timothy-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 184
- Figure 6.50 Numerical results for VTM-MRI-Bartholomew-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 185
- Figure 6.51 Numerical results for VTM-MRI-Barnaby-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 186
- Figure 6.52 Numerical results for VTM-MRI-Barnaby-/port/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 186
- Figure 6.53 Numerical results for VTM-MRI-Barnaby-/stern/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 187
- Figure 6.54 Numerical results for VTM-MRI-Barnaby-/food/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 187
- Figure 6.55 Numerical results for VTM-MRI-Barnaby-/neap/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae. 188

- Figure 6.56 Experimental results of MRI-based 3D-printed VT of Barnaby, phonating on /neap/ with (greyscale) and without (blue) vallecula. 189
- Figure 6.57 Experimental results of MRI-based 3D-printed VT of Barnaby, phonating on /neap/ with (greyscale) and without (blue) vallecula. 189
- Figure 6.58 Numerical (blue) simulation versus experimental (red) results of Barnaby singing on /port/. 190
- Figure 6.59 Numerical (blue) simulation versus experimental (red) results of Barnaby singing on /food/. 191
- Figure 6.60 Numerical (blue) simulation versus experimental (red) results of Barnaby singing on /stern/. 191
- Figure 6.61 Gaussian function. 195
- Figure 6.62 New metric for the SFC. 196
- Figure 6.63 Benchmarking of the new metric against prototypical cases. The behaviour of the new metric is being scrutinised on prototypical cases made of 1, 2 or 3 peaks of a Gaussian form, whose parameters (height, width, bandwidth) are varied. 199
- Figure 6.64 Changing voice type, from Bass-Baritone (Barnaby) to Soprano (Sophy) and the corresponding SFC obtained with the new metric. 202
- Figure 6.65 VTM-MRI of the six professional singers phonating on /hard/. 203
- Figure 6.66 VTM-MRI of Barnaby singing on different vowels. 204
- Figure 6.67 Barnaby singing on different vowels, with the SFC represented by the new metric. 204
- Figure 6.68 Bartholomew singing on different registers on /hard/. 205
- Figure 6.69 Bartholomew singing in different registers, with the SFC represented by the new metric. 205

- Figure 6.70 Barnaby singing on /hard/ 207
- Figure 7.1 A schematic of the hypopharyngeal cavities model. From left to right, the transfer functions of the VTp, VTp+epilarynx, VTp+epilarynx+piriform fossae+epiglottic vallecula. 220
- Figure B.1 Boundary between two adjacent volume controls K and L. 233
- Figure C.1 A triangular finite element partition of a domain Ω into elements over which test functions are piecewise linear functions of the coordinates (x, y) . 240
- Figure C.2 Element topologies supported in Actran, from www.fft.be. 242
- Figure C.3 A global basis function ϕ_i for the spaces U^h and V^h , created by assembling element shape functions which have the node i in common. The basis function is defined everywhere in Ω but is nonzero only on the subdomains surrounding the node i . 243
- Figure D.1 NMR: processes at work. 246
- Figure D.2 http://www.nobelprize.org/nobel_prizes/medicine/laureates/2003/illpres/ Diagram of an MRI system. 249

LIST OF TABLES

Table 2.1	Frequency of the low singing formant, frequency of the high singing formant, and the length of the vocal tract for different voice categories, from [25]	54
Table 2.2	Age, classification according to the Bunch and Chapman criteria [11], range, sung pitch (between brackets for falsetto), vowel and protocol acquisition for the 6 professional singers.	60
Table 5.1	“York” protocol: technical data for the MRI data acquisition of Maristela, Bartholomew and Barnaby.	104
Table 5.2	“Freiburg” protocol: technical data for the MRI data acquisition of Marisa, Timothy and Sophy.	105
Table 5.3	Dimensions of the different 3Dprinted pieces used to be combined to model the Vocal Tract (VTM-1 and VTM-2).	114
Table 6.1	Sung pitch, corresponding f_0 , five first resonance frequencies (in Hz) across voice categories for VTM-MRI-/hard/, average of R3, R4 and R5 and their statistical distribution (average μ and standard deviation σ).	154

Table 6.2	Vocal Tract dimensions versus SFC prediction. Epilaryngeal Volume (cm^3), Epilaryngeal Cross-Section (cm^2), equivalent Epilaryngeal Radius (mm), Vocal Tract Length (mm), Oral Length (mm), Pharyngeal Length (mm), Epilaryngeal Length (mm), Pharyngeal Width (mm), Vocal Tract Volume (cm^3), Piriform fossae Volume (cm^3), % the ratio of PV/VTV (in %), SFC (pred) (in Hz) the SFC frequency centre predicted from the dimensions of the hypopharynx (see subsection 6.4.2.1) 155
Table 6.3	Resonance frequencies shift (in%) when the epilarynx is lengthened downwards by 10 mm. Note that Marisa's MRI data showed some artifacts and were therefore more difficult to segment accurately: this is why the data obtained by lengthening the segmented epilarynx are not reliable and therefore put in between brackets. 172
Table 6.4	L_s is the averaged length of the two piriform sinuses [mm], f , the antiresonance frequency generated by the sinuses [Hz], VTV is the volume of the Vocal Tract [cm^3], PV is the volume of the piriform fossae [cm^3], % is the ratio PV/VTV expressed in percentage. The corresponding statistical average (μ) and standard deviation (σ) are also given. 180
Table 6.5	Comparison between the experimental and numerical formant frequencies (in Hz) of Barnaby phonating on different vowels. 193
Table 6.6	Comparison between the SFC predicted (pred) from anatomical dimensions and the SFC obtained with the new metric (met). 200
Table 6.7	SFC: comparison between the experimental and numerical results based on the new metric. 201

Table B.1	Comparison between implicit and explicit schemes.	231
-----------	---	-----

ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. David Howard, my supervisor, for giving me the freedom in my research and guide me with his knowledge and passion for the singing voice science.

I am very grateful to my friend Michael Spyres, an operatic rising star, who enlightened me with our talks and inspired me to start this thesis.

I would like to thank all the present and former members of the Audiolab for providing such a great atmosphere at work. Both funny and serious talks around cakes contributed to insightful ideas.

I would like to thank the fellow singers of my ensemble, Vox Luminis, who shared with me their performing experience from the artistic end of the singing science.

And last but not least, this thesis would not have been possible without the strong support of my large family. So I would like to thank my parents and my six siblings.

DECLARATION

I hereby declare that this thesis is entirely my own work and all contributions from outside sources, through direct contact or publications, have been explicitly stated and referenced. This work has not previously been presented for any other award at this, or any other, University.

Part I

INTRODUCTION

1 | HYPOTHESIS

1.1 SINGULARITY OF THE SINGING VOICE

Although much research has been carried out on the subject, the human voice remains one of the most curious, mysterious and thrilling ways of conveying musical content.

The vocal tract (VT) is roughly 17-20 cm long for the male adult and 15-18 cm long for the female adult [102, 20, 119]. Regarding other instruments of a comparable size, the panel of possibilities in terms of range, dynamics, registers and spectra is much wider for the voice [99]. Moreover, it is the only 'embodied' instrument: the individual being the exciter of the vibrating body as well as the source of the radiated sound [99]. As a consequence, there is a discrepancy between how the inner and outer ear perceive the sound produced [102]. Emotions directly affect phonation since they are intertwined with breathing and voice production [102].

Another particularity of the human voice is the text sung along with the melody. It helps to convey more information about the message beyond the music, but it makes this instrument more complicated to analyse.

1.2 HOW THE VOICE WORKS

Like other wind instruments, the voice is produced by the air passing through a valve (the glottis) situated between an upstream tube (from the lungs through the trachea) and a downstream tube (the vocal

tract) [119]. The glottis is composed of vocal folds which repeatedly open and close, letting air jets flow out [102]. When the upstream pressure is higher than the downstream one, it forces the vocal folds to open, and when the air goes through the glottis, which has a smaller cross-section than the trachea, it is accelerated. Due to Bernoulli's effect, it creates a suction that pulls the vocal folds together [102, 119]. These antagonist movements repeat in the form of an oscillator, generating acoustic waves that propagate through the vocal tract.

The vocal folds and the glottis are in a plane which defines two acoustic spaces relevant to speech production, i. e. the subglottal and supraglottal vocal tracts [66]. The subglottal cavity shape remains relatively constant during speech, but the supraglottal cavity undergoes complex modifications which create the various patterns of airflow and acoustic resonance that produce speech/singing [66]. The supraglottal vocal tract, or more commonly the Vocal Tract (VT) plays the role of a filter, shaping the harmonic signal generated by the oscillating vocal folds with a transfer function related to its shape and dimensions [32, 102, 109, 99]. In the larynx, at the glottis, the regular opening-closing of the vocal folds produces a harmonic signal with a fundamental frequency f_0 and its multiples [32, 102]. The VT then acts as a filter, enhancing several frequency bands, whose peaks are called *formants* [32]. The frequency response of such a filter is highly dependent on the shape and dimensions of the vocal tract [32, 102, 109, 99]. The sound produced is then radiated from the lips to propagate into the open atmosphere.

1.3 HYPOTHESIS

"The hypopharyngeal cavities and the vallecula play a predominant spectral role in the singing voice"

Decomposition of Hypothesis

The hypopharyngeal cavities

The hypopharyngeal cavities consist of the laryngeal cavity and the two piriform fossae, both fitting posteriorly at the bottom of the pharynx [66], as detailed on Figs 1.1, 1.2b and 1.2a.

- The *laryngeal cavity* extends from the end of the trachea to the connection with the pharynx. The vocal folds, whose narrow triangular separation is called the glottis, divides the laryngeal cavity into two parts:
 1. The *supraglottal cavity* is situated above the vocal folds. It is a short basal segment of the vocal tract between the glottis and the bottom of the mesopharynx [48] and consists of:
 - The *laryngeal ventricles*, or Morgani sinuses, a bilateral recess between the vocal folds and the ventricular folds [48]
 - The *laryngeal vestibule*, a narrow tract above the laryngeal ventricles [48]
 2. The *infraglottic cavity* is situated below the vocal folds
- The *piriform sinuses*, or piriform fossae, are the "pear-shaped" cavities lateral to the laryngeal vestibule [48], as seen on Figs 1.1, 1.2b and 1.2a. It is a pair of bilateral tubes of a conical shape formed by the aryepiglottic folds and the lateral hypopharyngeal wall, located above the closed entrance of the oesophagus [48]. Physiologically, they serve as side branches which "cap-

ture" foreign bodies, sometimes food and mucuses [3]. Some fluids must reside in the piriform fossae between swallow events [66].

The valleculae

The valleculae epiglotticae are depressions (vallecula) just behind the root of the tongue; they constitute a borderline area between the pharynx and the larynx: these depressions serve as a temporary saliva reservoir to prevent from swallowing [10].

Predominant spectral role

These cavities largely influence the spectrum of the voice.

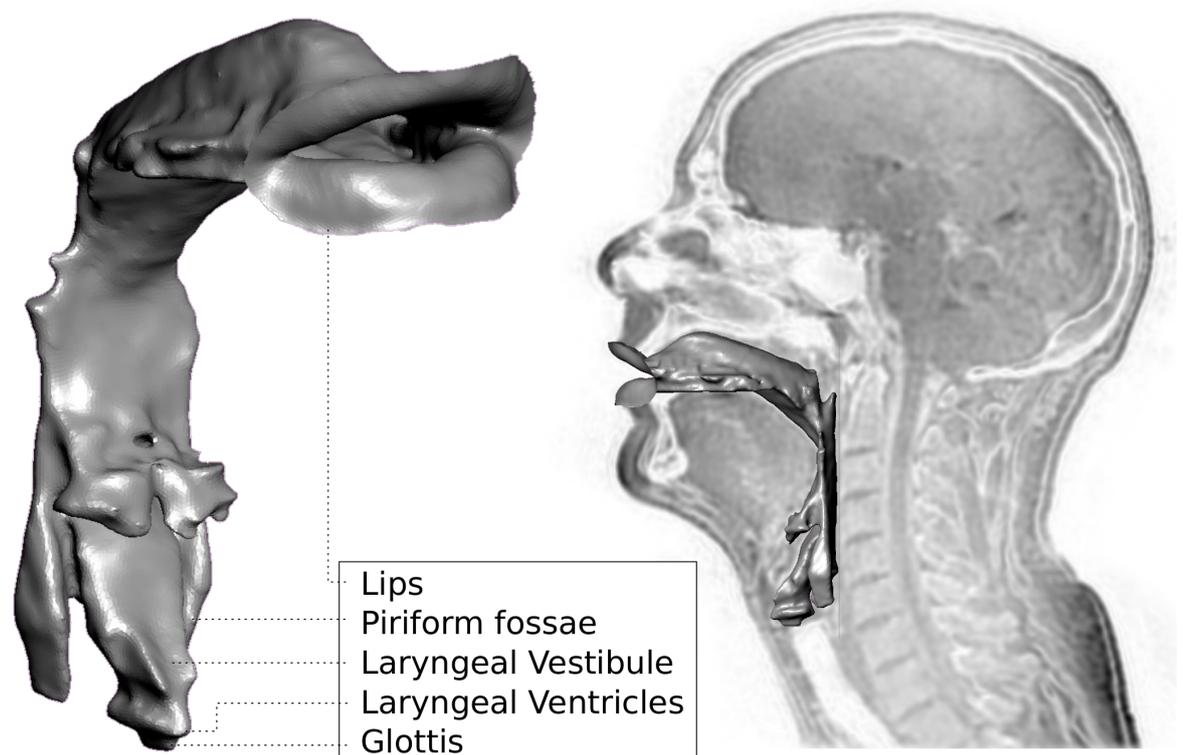
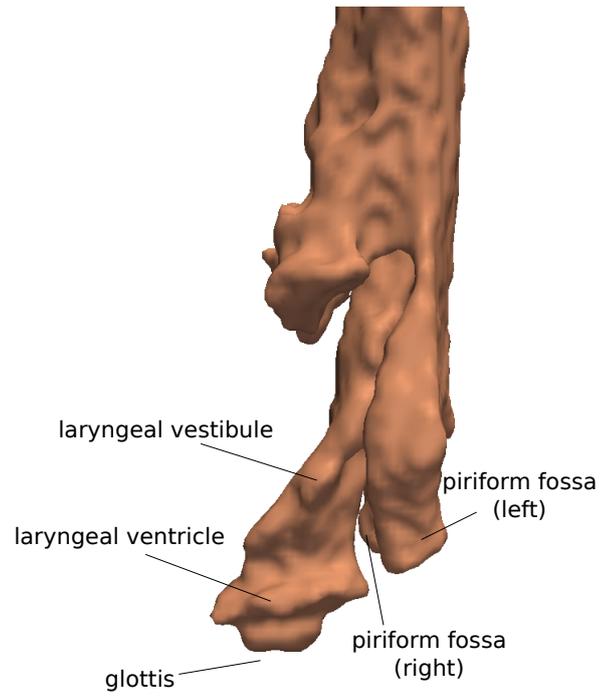
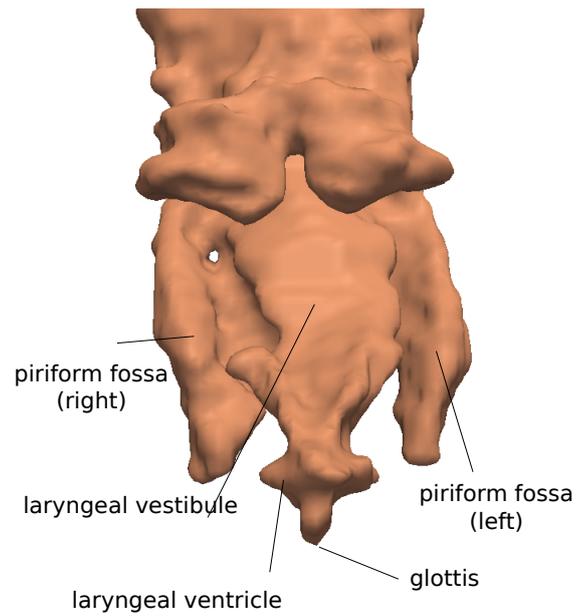


Figure 1.1: Vocal Tract and hypopharynx cavities superimposed on MRI mid-sagittal slice



(a) Lateral view.



(b) Frontal view.

Figure 1.2: Hypopharynx details.

1.4 THESIS OUTLINE

This thesis is organised as follows:

Part I: Introduction

Chapter 1 introduces the hypothesis and a brief summary of the voice production.

Chapter 2 gives an overview of singing voice production, followed by the Source-filter theory, focussing on the Vocal Tract and the origin of its resonances. This leads to an explanation of formants and the importance of the hypopharynx in relation to the Singer's Formant Cluster. The chapter concludes with an overview of the methods used to measure the Vocal Tract resonances and a summary of the results found in the literature regarding voice classification.

Part II: Materials and Methods

Chapter 3 explores the numerical techniques used to simulate the resonances of the Vocal Tract models. The chapter starts with an introduction of the Finite Volume Method in the time-domain, followed by the Finite Element Method in the frequency domain. To conclude, a note is given about the meshing which consists of dividing a 3D structure (here) into small linked volumes or elements linked to each other: the numerical methods need meshing as a preprocessing procedure.

Chapter 4 introduces a novel transducer-independent technique, adapted from the sine sweep method used mostly in room acoustics to measure the impulse response (and subsequently the transfer func-

tion) of the 3D-printed cavities (from simple models of the Vocal Tract to actual MRI-based 3D printed Vocal Tracts).

Chapter 5 describes the steps involved in the process from the acquisition of MRI data of professional singers phonating in a scanner to the 3D printing of their Vocal Tract.

Part III: Results

Chapter 6 introduces the theoretical predictions of the resonances of one tube and a twin-tube, emphasising the importance of the Open End Correction. The production of an extra resonance by appending one tube to another is introduced, as well as its consequences on the spectrum. The next section applies the findings to MRI-based Vocal Tracts of six professional singers and compares them with morphological dimensions. Eventually, it introduces a new metric for the Singer's Formant Cluster.

Part IV: Conclusion

Chapter 7 concludes and summarises the findings, suggesting areas for further development.

1.5 CONTRIBUTIONS

The novel contributions of the research carried out within the framework of this thesis are as follows:

- A novel transducer-independent technique, to measure the transfer functions of cavities, adapted from the method of Farina in room acoustics.
- Identification of the spectral impact of the epilaryngeal tube dimensions (length and radius) of Vocal Tract analogues.
- A tested new metric for the Singer's Formant Cluster.
- A formula for predicting the Singer's Formant Cluster related to the dimensions of the hypopharyngeal cavities.

In the framework of this research, I have:

- Confirmed the results found in the literature regarding the spectral impact of the piriform fossae and vallecula for MRI-based Vocal Tracts, both numerically and experimentally.
- Confirmed, for 6 professional singers, that the Singer's Formant Cluster predicted from their anatomical dimensions matches the one derived from their Vocal Tract resonances.

2 | SINGING VOICE PRODUCTION

Pioneers such as Fant [32] have investigated the basic principles of acoustic production in speech and singing, Sundberg [102] and Titze [109] explored the mechanisms at work in singing voice production and Stevens [89] focused on speech production. This chapter focuses mainly on the aspects of voice production relevant to this thesis. After an introduction about the speech/singing sounds and the anatomy of singing production, the source-filter theory is explained. The Vocal Tract (VT), acting as a filter, will then be given more attention, introducing the concept of formants and the so-called "Singer's Formant Cluster", thought to originate from the hypopharyngeal cavities dimensions. An overview of the methods used amongst the scientific community to measure the VT resonances follows, with a special development given to the VT shapes acquired by medical imaging. This chapter closes with a literature review on singing voice classification.

Speech/singing sounds

Speech/singing sounds are similar to any other sounds, in the way that they are microscopic and rapid fluctuations of air pressure around a reference pressure (usually the atmospheric pressure) [102]. Singing sounds are generated when the vocal organs set the air in movement. While speaking or singing, the acoustic energy is radiated through the air and when the pressure fluctuations reach the ear of the listener, the eardrum fluctuates in and out; in when the pressure is above the reference pressure, out when it is under [118]. The acoustical energy is then converted into mechanical energy, which will in turn be trans-

formed into neural energy so that the listener's brain can perceive and process the sound/voice they are hearing/listening to [113].

Speech waves

The sound is radiated from the lips and the speech/singing wave is distributed at any given instant, as an air pressure perturbation around the speaker/singer. A transducer placed in the radiating field emerging from the speaker/singer can, at any given point in space, transform the acoustical energy and convert it into electrical energy to record its fluctuations, as a time-pressure function, or a time-pressure wave [118].

Frequency

Looking closer at the waves recorded by the transducer, one can see rapid air pressure fluctuations around atmospheric pressure. These variations are periodic, with repetitive patterns, or oscillations. Measuring the time duration between two successive oscillations gives the period T of this oscillation. Taking the inverse of the period leads to the frequency $f = \frac{1}{T}$ expressed in s^{-1} or Hz [81]. For instance, the modern tuning $A = 440\text{Hz}$ given by the oboe before the start of a concert with orchestra to tune the instruments, or given by the tuning fork to set the pitch of a piece in an a cappella concert, means that the air pressure follows a cycle being above and then under the reference pressure 440 times per second [81].

Fundamental frequency

A transducer placed in the radiating field of a singer phonating on the pitch $A = 440\text{Hz}$ would record a pressure-time wave whose fluctuations would have different repetitive patterns. Measuring the time between two adjacent cycles of the longest repetitive pattern, and in-

verting it would give a frequency of 440Hz, called the *fundamental frequency*, noted f_0 [81]: this can be defined as the frequency at which the slowest pressure fluctuation of a complex sound oscillates [113]. In the case of a singer, phonating in normal conditions, i. e. no strohbass register or overtone singing, this corresponds to the rate at which the Vocal Folds (VFs) vibrate [113]. The temporal duration of each glottal flow pulse determines the fundamental frequency f_0 [102].

Partials

During phonation, the vibrating VFs do not only give rise to one single tone: instead, a series of tones or a *spectrum* is produced [102]. The lowest tone of this series is called the *fundamental* and the other tones are called the *overtones*. The fundamental plus the overtones form the *partials*: their frequencies form a *harmonic series* [102]. This means that the partial n has a frequency n times that of the fundamental. The frequencies of the partials are the multiple integers of the fundamental frequency f_0 [99, 102].

Fourier Analysis

The French mathematician Joseph Fourier developed the idea that any complex signal can be decomposed into an infinite sum of basic periodic signals, like \cos or \sin , for instance [88].

Spectrum

Taking the Fourier Transform of the sound speech/singing wave decomposes the signal in a series of frequencies (the fundamental and its harmonics) which are assigned a weight, or an amplitude, depending on their contribution in the speech/singing wave. There are several conventions to define the Fourier transform, such as in [88]:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x \xi} dx \quad (2.1)$$

where the independent variable x represents the time and the transform variable ξ the frequency.

Timbre

In the spectrum, several frequency bands are enhanced, others are decreased. The *spectral envelope* changes from one voice to the other and makes differentiation between two voices possible: this characteristic is called the *timbre* [102]. As will be seen in subsection 2.4, the vowel quality and voice colour (timbre) are determined by the shape of the VT [102, 99].

2.1 ANATOMY OF SINGING PRODUCTION

The anatomical parts involved in the production of speech and singing can be divided into two categories : the ones taking part in the *phonation* and the ones in the *articulation* [49]. The speech organs of phonation include the lungs and the larynx, whereas the ones of articulation include the cavities above the larynx plus the tongue, teeth and lips. The organs taking part in the speech/singing production are represented in Fig 2.1.

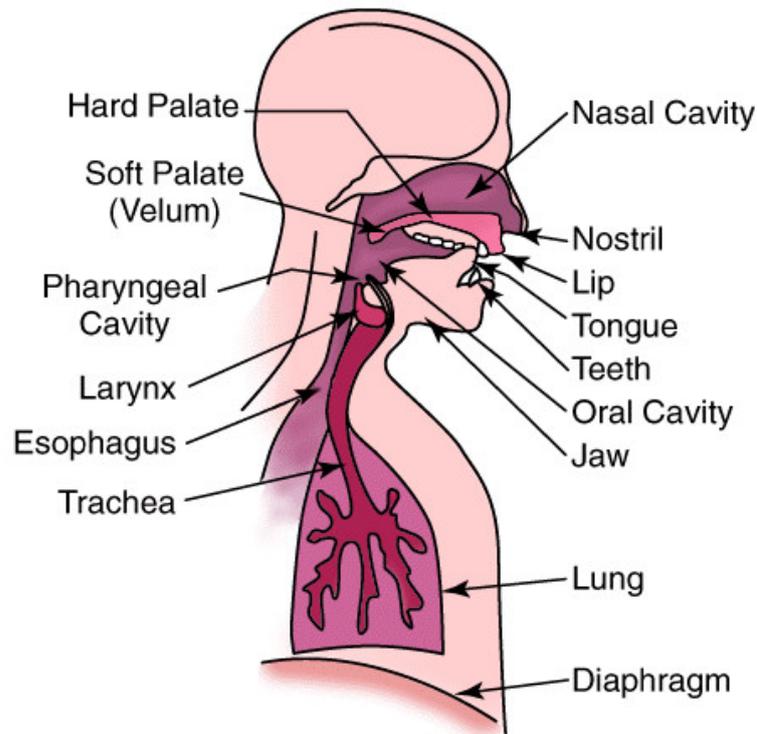


Figure 2.1: A sagittal view of the human speech/singing production, from [38]

2.1.1 Phonation

From the ancient Greek *φωνή* meaning voice, the term *phonation* relates to all the processes involved in the production of voice, whatever the form, i.e. laugh, cry, speech, singing, etc. The phonatory organs generate the acoustic source which will be modified by the *articulators* to produce the speech/singing. The first organs of phonation, the lungs, are also the largest: they act as a reservoir and provide the air necessary to create pressure fluctuations around the singer [113]. The diaphragm muscularly causes the lungs to inflate or deflate, expelling the air through the trachea to reach the larynx [102, 89].

The second organ of phonation, the larynx, which can be seen in Fig 2.2, hosts the Vocal Folds (VFs), which can be held shut through muscular coordination [66, 74]. The triangular opening made by the VFs during the opening phase is referred to as the Glottis [89]. Be-

sides its important role in phonation, from the evolutionary point of view, the larynx stands as a mechanism for fixation of the thoracic volume, a protective closure of the airway and expulsion of foreign particles trapped in the airway [66]. This allows closure of the respiratory system while swallowing food or liquid for instance [74]. When the VFs are sealed, as in effort closure, the larynx can be used to increase the abdominal pressure during physically demanding efforts such as weight lifting [66, 74].

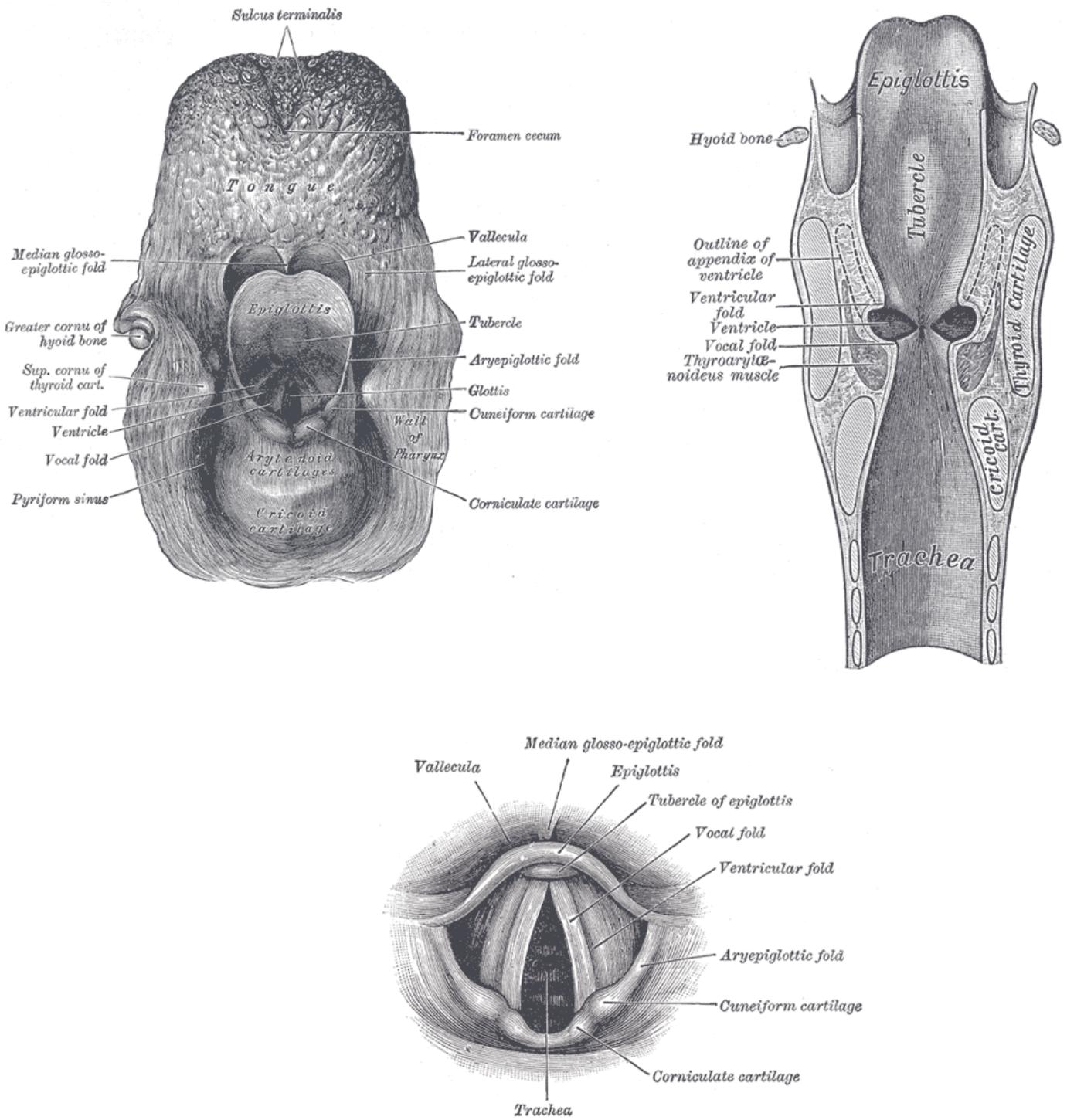


Figure 2.2: The entrance to the larynx, viewed from behind, Coronal section of larynx and upper part of trachea and Laryngoscopic view of interior of larynx. From Henry Gray (1825-1861). *Anatomy of the Human Body*. 1918. Figs 955, 954 and 956

In the phonation framework, the larynx is used, with the glottis opening and closing, as a generator of a quasi-periodic acoustic source, as can be seen in Fig 2.3. VFs are adducted during phonation, so that the space between them, the glottis, decreases. The air flowing through this constriction produces a pressure drop, resulting in a pressure excess from the lungs which tends to force the VFs apart and to accelerate air through the glottis [102, 110, 119]. The flow of air through the glottis creates a suction that tends to pull the folds back together ("Bernouilli effect", myoelastic-aerodynamic theory, Van Den Berg, 1958) [102, 119]. The myoelastic-aerodynamic theory (Van Den Berg, 1958) hypothesised that the vibration of the VFs is an interplay between these two forces [74], as can be seen in Fig 2.3.

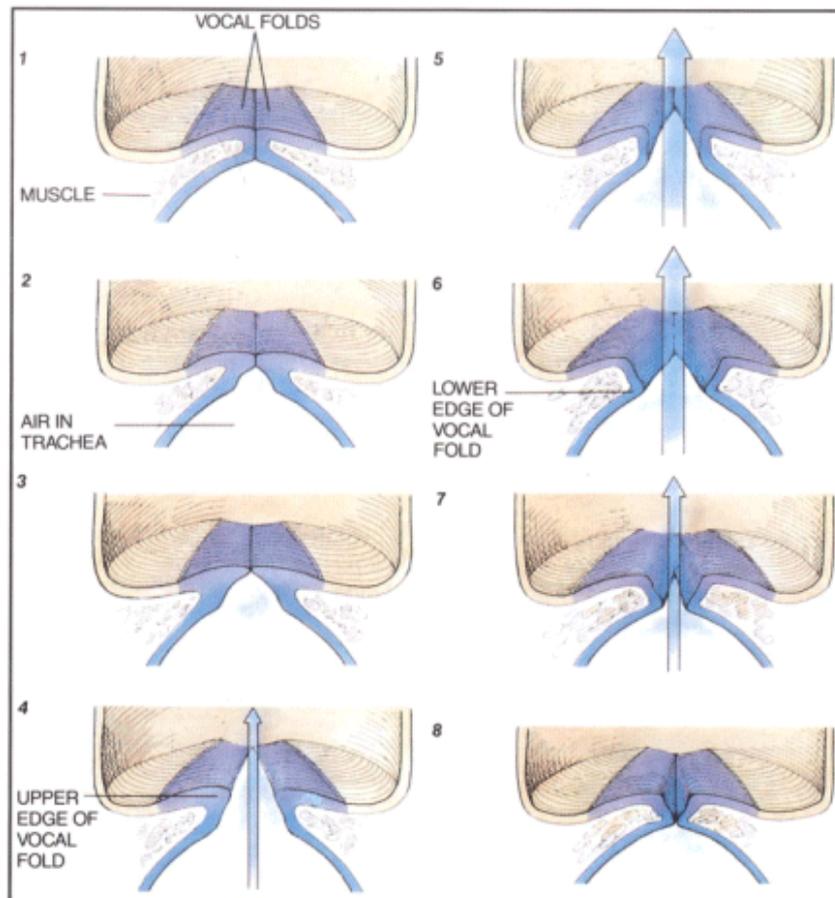


Figure 2.3: Vocal Folds vibratory cycle. The Vocal Folds are first held shut by muscular tension, but the increasing subglottal pressure eventually manages to open the sealed Vocal Folds. As air is passing through the glottis, the transglottal pressure decreases (Bernoulli's principle), causing a suction effect, causing the sealing of the Vocal Folds [102]. When the glottis is closed, the subglottal pressure is greater than the epiglottal pressure: if the vocal folds are adjusted for phonation, they cannot resist this gradient of pressure and open again [102].

The vibratory cycle of the Vocal Folds generates an acoustic signal, usually called the glottal flow signal [102, 89, 110, 74]. The air pressure and tension of the VFs determines their vibrating frequency, as in singing [102]. The vibrating Vocal Folds (VFs), by opening and closing periodically, generate airflow pulses in the Vocal Tract [99]. The VFs convert a quasi-steady air flow from the lungs to an oscillating air flow in the VT: in electrical analogy, this would mean that

the VFs convert a DC flow into an AC flow [119]. The VFs open and close at a certain rate, called the *fundamental frequency* f_0 . The integer multiples of this fundamental frequency, called the *harmonics*, $f_2 = 2 \cdot f_0, \dots, f_n = n \cdot f_0$ are excited as well, so that the source (vibrating VF in the larynx) generates a harmonic spectrum (f_0 + its harmonics, of an exponentially decaying amplitude) [99, 102, 58]. Muscles do not directly generate the vocal folds vibration, which is a passive aeromechanical effect (Van Den Berg, 1958), but contribute to its control [109].

2.1.2 Articulation

Articulation encompasses all the manoeuvres that change the vocal tract shape [102]. The puffs of air expelled out of the Glottis travel through the Vocal Tract (VT), made of the larynx, the pharynx, the oral cavity and the nasal cavity to radiate through the lips and the nostrils (see Fig 2.4) [38]. The *articulators* such as the pharynx, the tongue, the jaw opening, the soft palate (velum) and the lips can modify the geometry of the VT to shape the sound [102].

The Vocal Tract is a single multi-chambered tube starting at the glottis and ending at the lips, but for nasal sounds like /m/ or /n/, the nasal cavity acts as a parallel chamber to the oral cavity, creating two radiating ends, i. e. the nostrils and the lips respectively [74]. Fig 2.4 shows a schematic of the voice production.

2.2 SOURCE-FILTER THEORY

Around 1960, the Swedish scientist Gunnar Fant originated the idea that speech production could be explained by the Source-Filter theory [32]: it describes speech production as a filter plugged into a source,

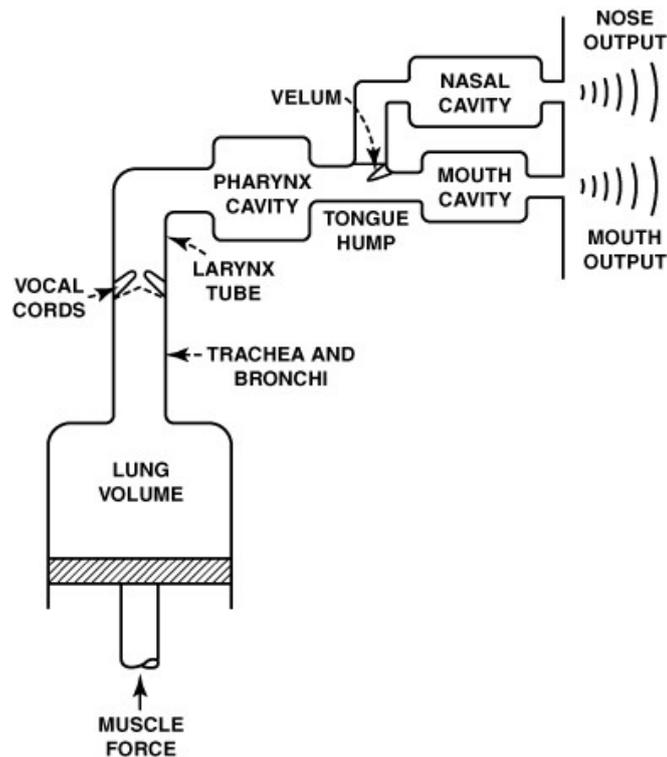


Figure 2.4: Schematic of the Speech/Singing production system, from [38]

both being independent of one another. The division between source and filter was described in the previous section, i. e. the phonation plays the role of the source whereas the articulation holds for the filter [102, 74]. The basic principle is that the acoustic signal generated at the Vocal Folds is shaped by the resonances and anti-resonances of the Vocal Tract [99], acting as a filter, and then radiated at the lips (and the nostrils) to the environment.

2.2.1 Framework

In the Source-Filter theory, the interactions between the filter and the source are neglected, i. e. it is considered that the sound waves in the Vocal Tract do not have an influence on the Vocal Folds vibration and vice-versa [32]. This hypothesis is valid as long as the fundamental frequency produced by the Vocal Folds is low in comparison with the first resonance of the Vocal Tract (see section 2.3, page 22).

2.2.2 The model

In Fig 2.5, the vibrating VFs are shown to produce the glottal flow spectrum at the level of the Glottis, whether it is a discrete spectrum (on the left), made of the fundamental frequency and its harmonics (in singing) or a continuous spectrum (on the right) in whispering. The continuous transfer function of the Vocal Tract is then superimposed to the glottal source spectrum, with its resonances and anti-resonances, as well as the radiation impedance spectrum which has a positive slope: high frequencies are better radiated at the lips [119]. A schematic of the resulting output sound spectrum is plotted at the bottom of Fig 2.5, at the left in singing, at the right in whispering. Since the vertical axis of the plots are logarithmic, a simple addition of the spectra gives the resulting spectrum.

The output spectrum contains the fundamental frequency plus the harmonics (forming together the partials), but their respective amplitude has been modified by the vocal tract resonances: harmonics near resonance frequencies are enhanced in amplitude whereas those distant from resonance frequencies are reduced [99]. Therefore, the glottal flow spectrum samples the frequency response of the Vocal Tract to express, via the sound output, information about both the harmonic content of the source and the resonances related to the shape of the Vocal Tract [99].

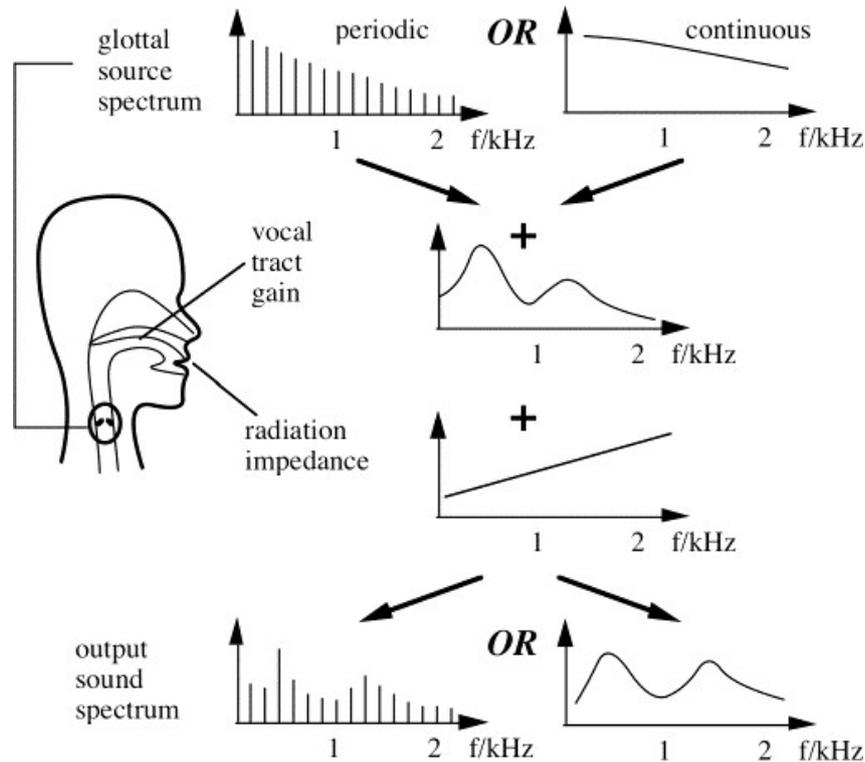


Figure 2.5: A schematic of the source-filter model, after [28]. The glottal source spectrum is produced at the glottis, whether it is periodic (singing, upper left) or continuous (whispering, upper right). The continuous transfer function of the vocal tract, as well as the positive radiation impedance spectrum are superimposed to the glottal source spectrum to form the output sound spectrum, either in singing (bottom left) or in whispering (bottom right).

2.3 VOCAL TRACT

"The upper part of the larynx, together with the pharynx, nares, and mouth, constitutes a passage-way, or tube, of variable size and shape, through which the vibrating current of air is passed. It is here that the voice is moulded, so to speak, on its way to the ear, and the shape of the passage-way largely determines the quality or timbre of the voice."

A.G. Bell, Mechanism of Speech, 1910 [7].

As mentioned previously, speech/singing is produced when air goes from an upstream duct (the trachea) to a downstream duct (the vocal tract), passing through a valve (glottis + vocal folds). The downstream duct (the vocal tract) has several strong resonances which can be modified by changing its geometric shapes [119].

The influence of the duct on the source depends mainly on the frequency of the first resonance of the duct. When the frequency of the valve f_0 is low in comparison with the first resonance of the tract (as is the case with most low pitched speech/singing), the duct has little or no influence on the fundamental frequency (f_0) but the geometry of the duct sets the threshold pressure for the oscillation [39]. The first resonance of the tract typically sits between 300 and 800 Hz [119]. In male singing, the f_0 of speech/singing is usually situated below this frequency, in which case the tract has little or no influence on f_0 [119]. However, the harmonics may fall near resonances or anti-resonances. In "normal" use of the voice (i. e. not for high-pitch use), it is therefore the tract resonances that have a great influence on the spectral envelope of the output sound rather than f_0 [119]. It is hence of importance to know how these resonances are created.

2.3.1 Origin the Vocal Tract resonances

The VT is roughly 15-20 cm long from the lips to the glottis and its resonances depend upon the length and the local constrictions of the tract [102, 20, 119]. The pressure wave, originated at the glottis, propagates through the airspace formed by the relative positions of the tongue, jaw, lips, and velum, and convey information about the shape of the tract which is eventually carried to the listener's ears [99]. This airspace, from the glottis to the lips/nares aperture, is called the *vocal tract*: it generates the acoustic characteristics of the sound output produced by a singer [99].

In their review, Wolfe et al. [119] develop extensively the notions of resonance and impedances. The present section is largely inspired by this review.

To understand the origins of the VT resonances, it is necessary to introduce the notion of acoustic impedance:

$$Z = \frac{p}{U} \quad (2.2)$$

which is the ratio between the acoustic pressure p and the acoustic flow velocity U [32, 109, 89]. Note that the acoustic pressure represents the variation of pressure from the steady reference pressure (such as the atmospheric pressure, for instance) and that the flow velocity is measured for a given section, e. g. at the glottis. Z is a complex quantity which varies with frequency. It takes different values along the length of the vocal tract. Its real part expresses pressure and flow in phase, whereas its imaginary part represents the flow and pressure 90° out of phase [119].

The real part dissipates acoustic energy (often in heat, through viscous friction) whereas the imaginary part accumulates acoustic energy [119]. A positive imaginary component occurs when the pressure is ahead of the flow. It is an inertive impedance, because it is related to the inertia of the mass of air: it needs a pressure difference to be accelerated and the acoustic energy is stored in its kinetic energy [119]. A negative imaginary component occurs when the flow is ahead of the pressure. It is a compliant impedance and the acoustic energy is stored by compressing and dilating the mass of air [119]. It can be compared with the well-known case of a spring oscillating around its equilibrium position. The mechanical energy oscillates between the kinetic and potential energy. The positive imaginary part of the acoustic impedance is like the kinetic energy of the spring whereas the negative imaginary part is like the potential energy of the spring. On the other hand, the real part is like the losses generated by friction from the spring and a surface or the air. The sign

of the impedance measured along a tract changes at each resonance [119].

At the lips, the sound coming from the VT sees the radiation impedance Z_{rad} , which is the pressure p needed to accelerate the mass of air just outside the mouth and produce an acoustic flow U_{rad} . This mass is small (cf. the Acoustical Length in 6.1, page 120). Consequently, Z_{rad} is small. From (6.4) page 128, it follows that the end correction decreases when the frequency increases. Hence, increasing the frequency decreases the mass of air to be accelerated outside the mouth. The pressure difference needed to set this mass of air in motion is smaller and therefore the radiation impedance decreases with frequency [119, 58]. Conversely, the transfer function increases with frequency making it easier for high frequency to propagate, hence the spectrum in Fig 2.5.

In the Vocal Tract as well as in a duct of a wind instrument, the impedance highly depends on the reflections of the sound waves. One of the strongest reflections happens at the lips, as the sound wave travels from a relatively high impedance inside the VT to a small radiation impedance [119]. If a high pressure burst is emitted at the glottis at the same exact moment that a high pressure burst, coming back from the lips, reaches the glottis, both their amplitude add up, generating a high impedance. Conversely, if a high suction burst, coming from the lips, reaches the glottis at the same moment of the emission of a high pressure burst, they (almost) annihilate and generate a small impedance. Resonances occur when the glottal flow creates large changes in flow or pressure at the lips [119]. These are happening at the minima of Z seen from the glottis (see Fig 2.7).

The most simple approximation of a Vocal Tract consists of an open cylinder, i. e. opened at the lips and closed at the Glottis. If the cylinder is of length L , the 3 first maxima of the transfer function

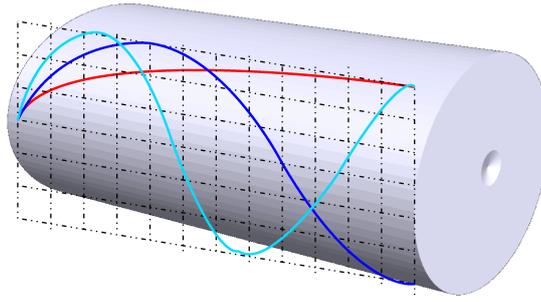


Figure 2.6: Pressure of the 3 first modes of an open cylinder, i. e. for $\lambda_1 = 4L$ ($\Rightarrow L = \lambda_1/4$ in red), $\lambda_3 = 4L/3$ ($\Rightarrow L = 3\lambda_3/4$ in blue) and $\lambda_5 = 4L/5$ ($\Rightarrow L = 5\lambda_5/4$ in cyan).

occur for wavelengths of approximately $\lambda_1 = 4L$, $\lambda_3 = 4L/3$, $\lambda_5 = 4L/5$, ... or for frequencies $f_1 = c/4L$, $f_3 = 3c/4L = 3f_1$, $f_5 = 5c/4L = 5f_1$, ..., as shown in Fig 2.6. Adding a narrow constricted opening at the glottis, this cylinder roughly behaves like the VT of a subject phonating on the mid-central vowel "schwa". Fig 2.7, taken from [119], shows the difference between an open glottis and a narrow constricted opening at the glottis. In the ideal case, the minima of Z occur between the maxima of Z , but appending a narrow constricted glottis moves the minima of Z (pressure node, flow antinode) towards lower frequencies. For a sufficiently small glottis, the slopes in $Z(f)$ are almost vertical: the maxima and minima of the impedance function happen at very close frequencies, and so do the maxima in the transfer function, which are the resonances of the tract.

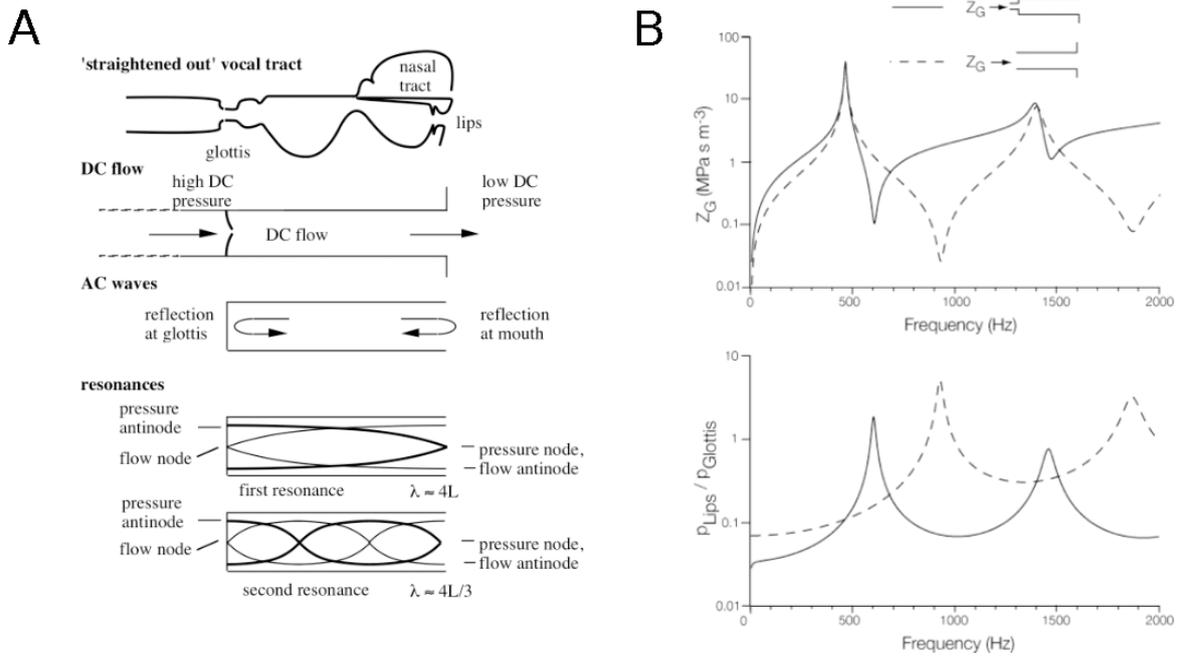


Figure 2.7: Figure taken from [119]. Part A shows successively a straightened Vocal Tract, emphasising the location of the different constrictions from the glottis to the lips, The conversion between a DC flow coming from the trachea and the lungs to an AC flow in the Vocal Tract, with successive reflections at the mouth and at the glottis, and then the first resonances of an open tube (closed at the glottis and opened at the lips). Part B gives the impedance and transfer function of a cylindrical vocal tract of length=170 mm and radius=15 mm with and without an appended narrow constricted glottis.

2.4 FORMANTS

The term "formant" has been used to designate different things across disciplines. Originally, Fant [32] gave the definition: "*The spectral peaks of the sound spectrum $|P(f)|$ are called formants.*" He defines resonances as the peaks of the gain function of the tract $|T(f)|$: "*The frequency location of a maximum in $|T(f)|$, i. e., the resonance frequency, is very close to the corresponding maximum in spectrum $P(f)$ of the complete sound*", adding

"Conceptually these should be held apart but in most instances resonance frequency and formant frequency may be used synonymously." Wolfe et al. [119] observe that for some voice researchers, the formant is a peak in the sound spectrum (a property of the speech/singing sound), for others it is a resonance of the vocal tract (a physical property of the tract), while a third group thinks of it as the pole in a mathematical filter model of the vocal tract (a property of a model). In this thesis, both the concepts of VT resonances, R_1, R_2, \dots, R_i and formants F_1, F_2, \dots, F_i as peaks in the sound output are used.

2.4.1 Consequences for the human voice

The acoustic energy generated at the glottis is more effectively radiated near the resonances of the tract, giving rise to broad peaks in the output sound spectrum, the formants [119]. The Vocal Tract (VT) of a singer acts as a filter on the acoustic output from the vibrating vocal folds, enhancing some of the harmonics of the glottal signal [102, 119, 99]. The formants F_1 and F_2 (and to a lesser extent F_3) are responsible for the vowel identification [78, 72, 102, 47, 58] whereas higher formants (F_3, F_4, F_5, \dots) relate to the voice quality [99], or timbre [119, 58], or *tone colour* [32, 102]. Lowering the velum (soft palate) couples the nasal tract with the oral tract. Not only does this allow for the production of nasal consonants and vowels, but it modifies the frequency and amplitude of the oral resonances, subsequently changing the behaviour of the formants of the sound output [36, 13].

There are many possibilities to use the previously mentioned articulators to act on the resonances of the tract(s), but the Vocal Tract also includes some regions, like the hypopharynx, whose geometry are almost fixed, which would result in an individual speaker's timbre of voice [57].

2.4.2 The Singer's Formant Cluster

The overall voice quality can be significantly influenced by the relative proximity of the resonances R3, R4 and R5 of the vocal tract: it is in this region that the so-called "singer's formant", or "singing formant", or "singer's formant cluster" appears [99]. An example can be seen in Fig 2.8, where the LTAS of a tenor shows a broad peak of spectral energy in the region around 3000 Hz, which is generally produced by a cluster of two or more resonances of the vocal tract, whose positions are close to each other, according to Sundberg [101, 102].

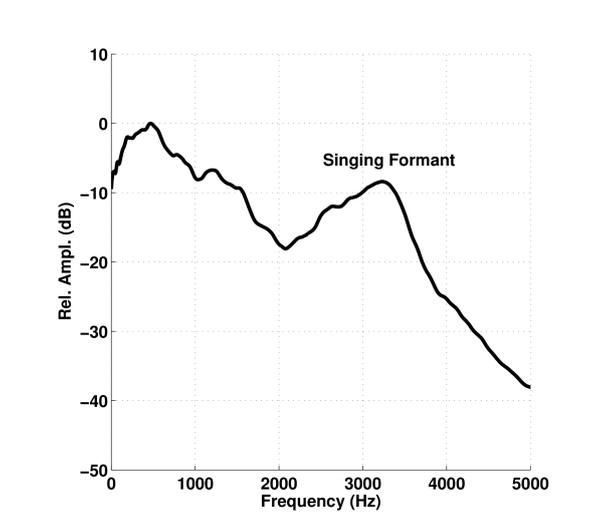


Figure 2.8: Long time average spectrum (LTAS) of a tenor. A singing formant is apparent in the range of 2500-3500 Hz (from [99]).

Bartholomew [6] seems to have been the first to highlight the presence of a broad peak around 2800-2900 Hz in the spectrum of male operatic singers. He termed it the "high formant" and suggested that it exists "regardless of whether produced by a tenor or a baritone,... and regardless of fundamental pitch [f_0], the vowel or intensity". Bartholomew narrowed the origin of this "high formant" to the part of the vocal tract between the "rima glottidis [glottis]" and the top rim of the "laryngo-epiglottal funnel", or essentially what is now called the epilarynx (see 2.4.3.1). Lewis [62] questioned the theory

that the singing formant was issued by a fixed resonator. He performed a spectral analysis on several male singers. His data showed a spectral prominence around 2800-3200 Hz but he suggested that the frequency location was dependent on the singer and the vowel being sung. Chiba and Kajiyama [14] regarded the formant around 2800-3200 Hz observed by Lewis as the resonance frequency of the laryngeal cavity, because the frequency matched with the resonance frequency computed from an effective laryngeal cavity of a tube length of 2.8 cm.

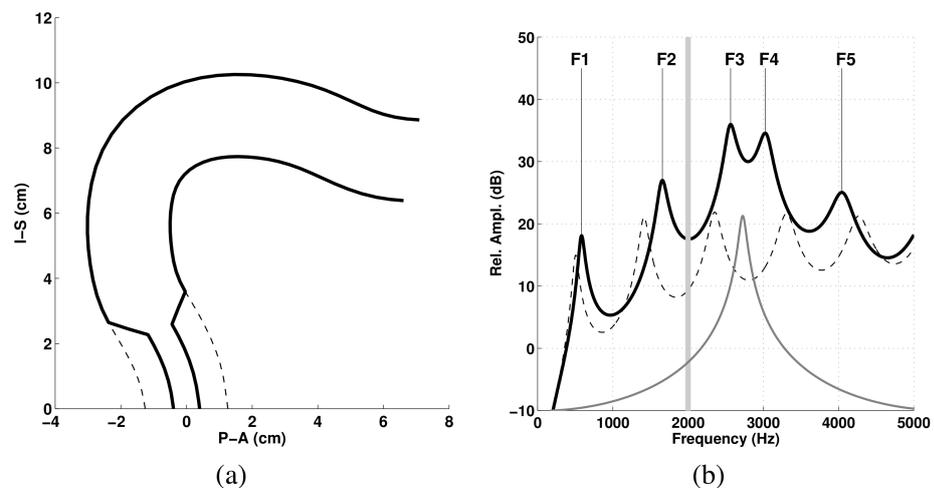


Figure 2.9: Idealised vocal tract modification to generate a singing formant.

(a) Uniform tube (dashed) and the uniform tube with a narrowed epilarynx (solid). (b) Frequency response functions of both the vocal tract shapes in (a) using the same respective line styles; the gray line represents the frequency response of the epilaryngeal tube in isolation. Figure taken from [99].

Originally, Sundberg [101] described the singer's formant as a spectral envelope peak around 2-4 kHz, typically observed in adult male western opera singers, which is independent of the vowel being sung: he suggested that the "singer's formant" was generated by the epilarynx tube which can act as an independent resonator as long as the ratio of its cross-section to that of the pharyngeal entry is equal to or less than 1:6. The singer has control of the epilarynx tube by lowering

the larynx and/or constricting the epiglottal and lower pharyngeal region [99]. Story shows the effect of an epilaryngeal tube resonator in Fig 2.9: a pseudo-midsagittal plot of an idealised VT configured first as a uniform tube (dashed lines) with a 5 cm^2 cross section then with the epilaryngeal portion constricted to a 0.5 cm^2 (solid lines) [99]. The frequency response is shown on the right panel of Fig 2.9, where the constricted epilarynx generates a spectral prominence by moving F_3 and F_4 towards each other. The newly acquired proximity of F_3 and F_4 causes their respective filter skirts to overlap and create a cumulatively greater amplitude response than when F_3 and F_4 were further apart (in the original setting of the uniform tube) [99]. The grey line shows the frequency response of the standalone epilarynx tube, whose resonance frequency can be computed with the formula of a closed-open tube, $f = c/4L$ [99]. Therefore, conceptually, the epilarynx resonance frequency acts as a "formant attractor" [99, 108].

A narrowed epilarynx is a possible mechanism to elicit the creation of a singer's formant: measurements of this narrowing have been reported in numerous studies [91, 93, 96, 96, 98, 108, 110, 27, 101, 50, 18, 57, 104, 105]. However, the existence of the singer's formant without an apparent lengthening and narrowing of the epilarynx tube has been reported by other researchers. Detweiler [23], using magnetic resonance imaging (MRI), stroboscopy, and acoustic analysis, found that the singers in her study produced a spectral prominence around 2800-3200 Hz, but did not show a 1:6 ratio of cross-sectional area of the epilarynx to the entry of the pharynx. Moreover, Wang [116] reported that the singer's formant can be achieved with both a low and a high larynx. Both studies suggest that the singer's formant can be produced by vocal tract modifications other than lengthening or narrowing the epilarynx.

The acoustic power of an orchestra steadily declines with frequency above a couple of hundred hertz, giving the opportunity for a voice exhibiting the "singer's formant" to be heard over a large or-

chestra in a big opera theatre or concert hall without any electronic amplification [119]. The "singer's formant" is now termed "singer's formant cluster" (SFC) due to its manifestation through a cluster of the third, fourth and/or fifth resonance of the tract [101, 102, 119]. If two or more of these resonances are close to one another, they can produce a single broad peak in the spectrum [119, 99].

Based on small-perturbation analyses, of area functions, Sundberg [103] reported that F_4 , in particular, but also F_5 of the sung vowel /a/ and /i/, is highly sensitive to the area function of the laryngeal cavity, although other regions of the area function also affect the locations of these formants.

Due to wide harmonic spacing, the SFC in the voices of women, especially (high) sopranos, either seems to be weaker, not well documented or non-existent [117, 114]. The previously mentioned use of the singer's formant does not appear to be a necessary or useful condition at high f_0 's to be heard over an orchestra. Indeed, in these cases, the fundamental frequency can be far above the threshold frequency from which the acoustical power of the orchestra steadily declines [119]. Moreover, high voices take advantage of resonance tuning strategy, as explained below [45].

Combining the high harmonics produced by the rapid closure of the glottis with a strong SFC enhances the energy radiated in the range 2-4 kHz, which corresponds both to the frequency region to which the human hearing apparatus is most sensitive [39, 48] and where the spectral power of the orchestra has drastically fallen [119]. This appears to be a reason why professional singers use the SFC to be able to perform in large venues without amplification [119]. Another advantage of radiating more energy in the range 2-4 kHz resides in the directivity of these frequencies. Indeed, at low frequencies, the wave-length is big in comparison with the VT dimensions and it radiates rather omnidirectionally, whereas in the 2-4 kHz

range, the wavelength is of the order of $\lambda = c/f = 340\text{ms}^{-1}/2000\text{Hz} = 17\text{cm}$ to $\lambda = c/f = 340\text{ms}^{-1}/4000\text{Hz} = 8.5\text{cm}$, which corresponds to the length of the VT or smaller [102, 119]. In that case, it radiates mostly in the direction of the mouth opening. Since performers usually sing facing the audience [38, 55], the use of the SFC can be seen as a means to not "waste" acoustic energy radiating in all directions by focusing a part of this energy towards the audience, who will eventually receive the acoustical energy [119].

A NOTE ABOUT RESONANCE TUNING

High voices, such as sopranos, typically sing in a range in which fundamental frequencies overlap with the range of the first resonance of the tract, i. e. 300-800 Hz. When they increase f_0 and try to maintain typical resonance values associated with particular vowels, two problems occur. Firstly, for some vowel-pitch combinations, f_0 would fall above R1 or even R2, losing the benefit of impedance matching of the first and/or second resonance, which allows the singer to produce sustained singing with relatively small vocal effort [119]. Secondly, the sound output would exhibit large variations across vowel-pitch combinations, making the voice highly inhomogeneous [119]. It is therefore sometimes necessary in singing to modify a vowel in order to achieve desired production of a certain note [110]. If a musical score, for instance, requires the note F4 = 349Hz to be sung on an /i/ vowel, a singer could not use the /i/ vocal tract shape without sacrificing vowel identity and possibly stability of phonation [108]. This is due to the fact that R1 of the vocal tract configuration of /i/ is around 300 Hz, which is below f_0 . The solution is to slightly modify the /i/ shape so that R1 is increased to be slightly higher than f_0 [99].

A number of studies suggest singers adopt what is termed "resonance tuning strategy" [53]: they adjust the resonances of the tract in order that the resonances such as R1 and/or R2 match the fundamental frequency f_0 or its multiples ($2f_0, 3f_0, \dots$). Joliveau et al.

[53] and Bresch et al. [9] showed that at low pitches, sopranos sing with typical R1 and R2, but when f_0 rises, they adjust the value of R1, so that R1 falls slightly above f_0 , as seen in Fig 2.10 from [53]. This results in two effects: above a certain pitch, the vowel is altered and the voice is more homogeneous for any vowel-pitch combination. Henrich et al. [45] showed that voice types other than sopranos are also concerned with resonance tuning. They measured R1 and R2 for 22 classically trained sopranos, altos, tenors, and baritones while they were singing four different vowels over their normal pitch range, by injecting a broadband acoustic current into their tract while phonating. The results indicated that resonance tuning was used across all voice ranges, and that different resonance tuning strategies were used for lower voice types.

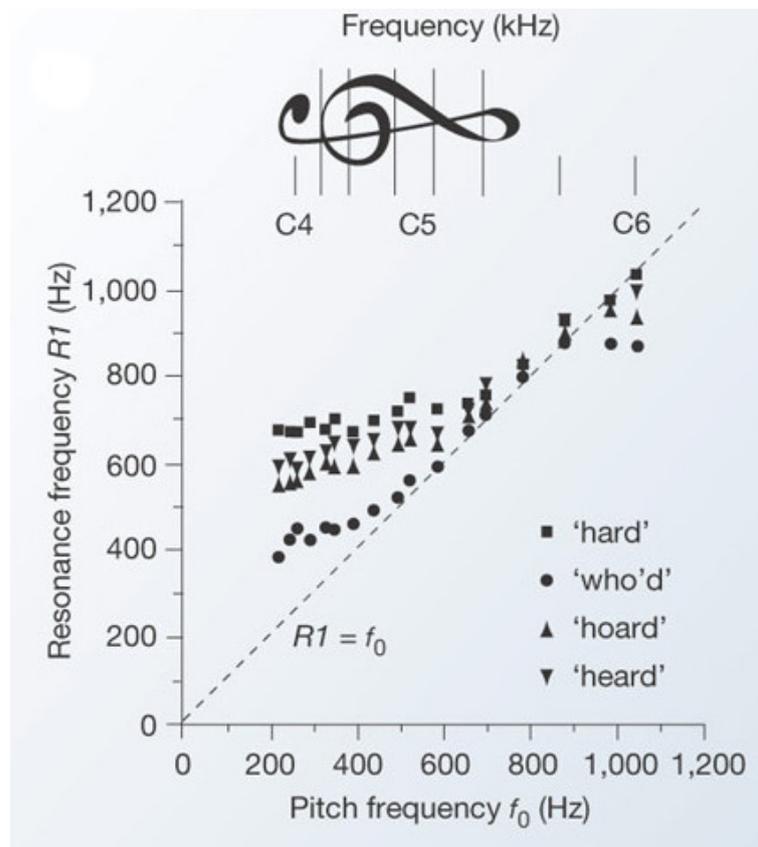


Figure 2.10: The fundamental frequency f_0 versus the first resonance R1 for sopranos. When f_0 is of equal or greater value than R1, sopranos tune R1 to a value slightly above f_0 . Taken from [53]

2.4.3 The Hypopharynx

The hypopharyngeal cavities consist of the laryngeal tube and the piriform fossae (see Figs 1.2a and 1.2b). The laryngeal tube (or epilarynx) is divided into the laryngeal vestibule and the laryngeal ventricles [57]. The laryngeal ventricles (or Morgagni's sinuses) correspond to a narrow bilateral hiatus between the vocal and vestibular folds and the laryngeal vestibule forms a narrow conduit from the ventricles to the opening into the mesopharynx [48]. The piriform fossae are bilateral cavities, located on each side of the laryngeal vestibule. These two cavities, shaped as an inverted cone, are opening into the lower part of the pharynx [57].

The hypopharynx is involved in vital human activities of phonation, respiration and deglutition [48]. The vocal and vestibular folds abduct to expand the laryngeal airway during respiration, while the same structures form the laryngeal ventricles and vestibule for voice production [48]. The piriform fossae contribute to the widening of the laryngeal airway, acting as a reservoir for full abduction of the arytenoid cartilages [48].

F₃, F₄ and F₅ are thought to relate to the geometry of the hypopharyngeal cavities [102], with large inter-speaker variations and small intra-speaker (i.e., inter-phoneme) variations [57]. The shape of the hypopharynx is relatively stable during the phonation of a vowel, regardless of vowel type, whereas the rest of the vocal tract (termed the Vocal Tract proper [104], VT_p in this thesis) shows large changes during vowel production [57, 67].

Kitamura et al. [57] were looking for a region in the vocal tract which would account for the speakers' acoustical individualities, with large inter-speaker variations and small intra-speaker (i. e. inter-phonemes) variations. One region potentially satisfying these conditions is the lower part of the vocal tract, i. e. the hypopharynx: Dang et al. [18]

observed that the area function (see section 2.5.1) of the piriform fossae is relatively stable during the phonation of different vowels with the same individual, whereas it exhibits large variations across individuals. Takemoto et al. [105] reported that the area function of the laryngeal vestibule was almost constant across vowels phonated by a male subject, obtained by a 3D cine-MRI technique.

Considering this potential area as the root for small intra-speaker and large inter-speaker variability respectively, Kitamura et al. [57] analysed MRI data obtained from four subjects and carried out morphological analysis on the mid-sagittal and transverse planes. Finally, they performed an acoustical simulation using a transmission line model to investigate the spectral intra-speaker / inter-speaker variability due to the hypopharyngeal cavities. Their results indicate that the hypopharynx shows relatively small intra-speaker variation and relatively large inter-speaker variation. They conclude by suggesting that the global shape of the vocal tract provides speaker individualities in the lower frequency region of speech spectra, while the hypopharynx provides those in the higher frequency region.

These results were supported later by Takemoto et al. [104] who reported that the volume from the glottis to the lower pharynx was almost constant (0.39 cm^3 standard deviation) during the utterance of a sequence of Japanese vowels /aiueo/.

Honda et al. [48] also found that the shape of the hypopharynx is relatively stable during natural vowel production regardless of vowel types, whereas the rest of the vocal tract region (vocal tract proper) shows large changes during vowel production [57, 67].

Based on the findings presented above, the hypopharyngeal cavities resonances play an important role in determining the voice quality, because it influences the frequency region to which human ears are most sensitive [39, 48] and it is engaged in vocal control for singing [48].

The singer's formant cluster (SFC) is a well-established feature of the acoustic output from the VT of trained opera singers that is independent of the vowel being sung [101]. It is commonly described as a cluster of F₃, F₄ and F₅. This suggests that it is related to a region of the VT that does not change greatly in shape with vowel articulation; anatomically, this relates to the hypopharyngeal cavities. Honda et al. [48] explains the creation of this broad peak with three actions:

1. Laryngeal manoeuvres to lower the peak frequency of the laryngeal cavity resonance, either by widening the laryngeal ventricles or by constricting the laryngeal vestibule.
2. Extra-laryngeal forces to widen the piriform fossae by advancing the laryngeal structures so as to deepen the antiresonance trough and lower the fifth resonance frequency.
3. Vocal Tract adjustment for raising the third formant frequency if it is necessary.

They add that *"These actions result in a relatively constant amplitude level of the formants up to the singing formant with a sharply declining spectral slope above the peak frequency of the singing formant."* They conclude with this paragraph:

The hypopharyngeal-cavity resonance is also important as one of the causal factors of individual vocal characteristics [...] individual differences in vocal-fold length and vocal-tract length are clear causal factors of speaker idiosyncrasy. [...] there is a frequency region that signals speaker-specific sound characteristics. [...] the hypopharyngeal cavity resonance contributes to realising the idiosyncratic nature of voice by determining the higher frequency spectra, presumably together with a speaker-specific distribution of vowel formants including a certain spectral interaction between the hypopharyngeal-cavity resonance and the higher formants from the vocal tract proper.

2.4.3.1 *The epilarynx*

The term *epilarynx* or *epilaryngeal tube* defines both the spaces it encloses and the physical structures it comprises [66]. Seeing the trachea and the larynx as a long tube, the epilarynx is the supraglottal, tube-shaped, upper extension of this tube, and is roughly 2 cm long for males [66]. In the framework of the vocal tract, the epilarynx is a tube that opens into the larger pharyngeal tube [101]. The vocal folds, the epilarynx and the pharynx together form the lower posterior region of the supraglottal vocal tract (or *lower vocal tract*), as illustrated in Fig 2.11: the epilarynx is presented as a "tube in a tube"; the upper part of the tube which fits into the inferior region of the pharynx [66]. It can be noticed from Fig 1.1, that the pharynx does not simply blend continually into the epilarynx, but prolongates laterally, downwards, to terminate at the bottom of the piriform fossae. Hence, it is referred to as a *tube within a tube* [66].

The *epilarynx* defines the set of structures comprised of the ventricular and aryepiglottic folds, the epiglottis and the arytenoid cartilage [66] (see Fig 2.12). It encloses two (acoustical) spaces as defined in Figs 2.12 and 2.13: the laryngeal ventricles and the laryngeal vestibule [66].

Physiologically, the role of the epilarynx reflects the evolutionary origin of the larynx as a mechanism for fixation of the thoracic volume, protective closure of the airway, and expulsion of foreign bodies trapped in the airway [66]. Due to the low position of the larynx relative to the velo-pharyngeal port, humans are potentially vulnerable to fatal aspiration of food or other substances, especially during swallowing; the epilarynx plays a role in the closure mechanism that protects us [66]. The closure of the epilarynx during throat clearing and coughing allows for the build-up of a sufficiently high intra-thoracic pressure; when the epilarynx opens, the abrupt high-

velocity of air expels any intrusive substance within the airway of the epilarynx [46]. In effort closure, such as coughing or heavy weight lifting, the bulging of the folds (both ventricular and vocal) become substantial to the extent that the entire epilaryngeal tube is narrowed and eventually closed off, as an effective hermetic seal [77]. As a protective function, the aryepiglottic folds and epiglottis are essential in preventing foreign matter to enter the laryngeal airway [66]. Moreover, the aryepiglottic folds create a steep wall which allows moderate pooling of fluid in the piriform fossae during a continuous manoeuvre of drinking [37]. Note that the ventricles secrete mucous providing lubricant for both the vocal folds and the larynx [66].

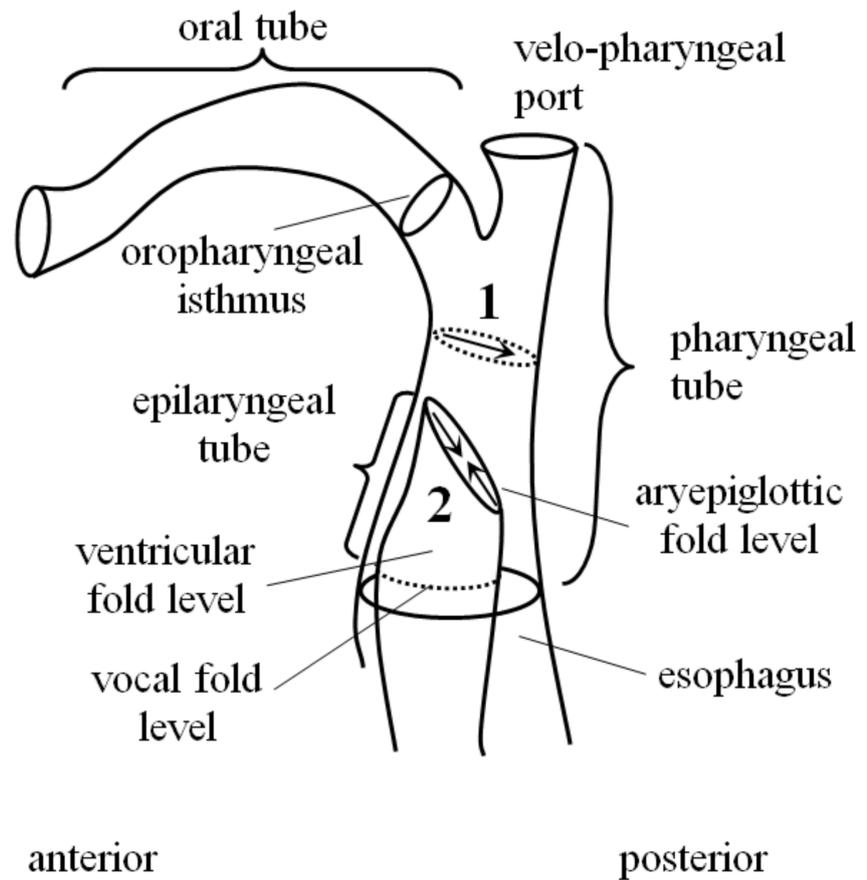


Figure 2.11: The epilarynx as a tube-in-a-tube. The tube-shaped epilarynx is found at the bottom of the pharynx tube; together these structures define the lower vocal tract. The action of these tubes are pharyngeal constriction (1) and epilaryngeal constriction (2). From [66].

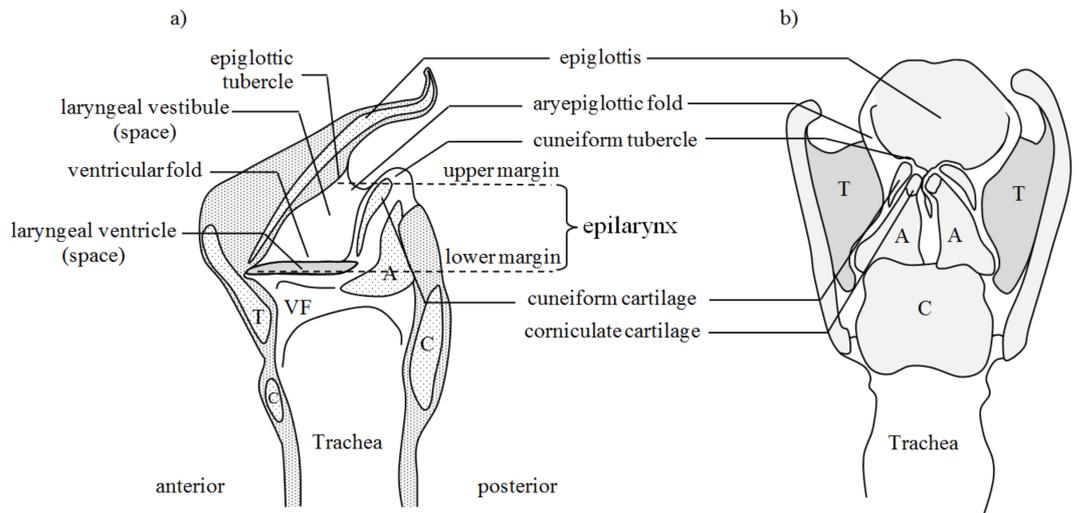


Figure 2.12: Anatomical sketch of the larynx. Sagittal section (a); posterior view (b). A = arytenoid cartilage; C = cricoid cartilage; T = thyroid cartilage; VF = vocal fold. Illustrations of laryngeal anatomy important to the epilaryngeal tube. From [66], after [122].

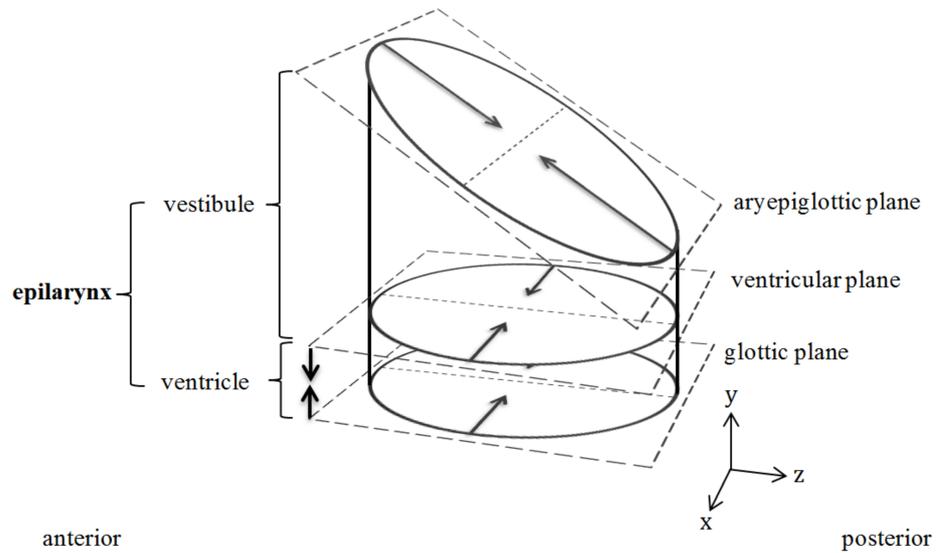


Figure 2.13: Functional planes of the larynx. Arrows indicate the axes of activity in each plane, the opposed vertical arrows between the ventricular and glottis plane indicate that these planes make contact. The dashed lines inscribed on the ellipses indicate lines of stricture. Dimension: x-axis is lateromedial; y-axis is inferio-superior; z-axis is posterior-anterior. From [66].

The laryngeal cavity, or epilarynx, is the region from the glottis to the junction at which its narrow exit is connected with the entrance of the wide pharyngeal cavity [104]. The epilarynx being of an intermediate diameter between the glottis opening and the pharynx, its characteristic impedance has an intermediate value, between the one at the glottis and the one at the entrance of the pharynx, turning the epilarynx tube into an impedance matcher at some frequencies [108, 112]. An impedance matcher is ideal to transmit acoustical power both downstream and upstream [119]. Downstream, it increases the transmission of the acoustic energy for some frequencies, enhancing some frequency bands of the sound output from the resonances of the tract. Upstream, it allows for a source-filter interaction, with pressure waves of the tract on the folds [119].

Fant [32] performed perturbation analysis on the six Russian vowels and found that the laryngeal cavity had a significant influence on the F_4 of /a/, /u/, /i/, and on F_5 for all the vowels except for /i/. Titze et al. [108] reported that a narrow epilarynx acts as a *formant attractor*, attracting the formants in the region 2.5-3 kHz, confirming the attraction of all the formants frequencies toward a single frequency focus observed by Sundberg [101]. Titze et al. add that the combination of a narrow epilarynx tube and a wide pharynx is ideal for maintaining a positive and steadily rising inertive reactance for $0 - F_{e1}$, at the expense of a more compliant reactance above F_{e1} , where F_{e1} is the first resonance frequency of the epilarynx, as a quarter-wave resonator. Fant and Båvegård [31] showed that the length of the laryngeal tube affects the frequencies of F_4 and F_5 , reporting that shortening the laryngeal cavity by 0.5 cm increases F_5 greatly and F_4 to a lesser extent. Titze [110] suggests that a narrow epilarynx tube helps in maintaining a highly inertive vocal tract, which is the key to a resonant voice. Takemoto et al. [105] found that the expansion (resp. constriction) of the laryngeal ventricle or constriction (resp. expansion) of the laryngeal vestibule reduced (resp. increased) the frequency of F_4 .

Takemoto et al. [104] studied the transfer functions of the whole vocal tract, the laryngeal cavity, and the vocal tract proper (VTp, the whole vocal tract deprived of its laryngeal cavity) from the area functions for the five Japanese vowels obtained from an adult male speaker and reported that the laryngeal cavity generates F₄ and that both the laryngeal cavity and the piriform fossae only slightly affect the other formants. Honda et al. [48] suggested that the epilarynx acts as a Helmholtz resonator, with wide ventricles (cavity) and narrow vestibule (neck), to produce a spectral peak around 3-3.5 kHz: in the vowel spectra, this resonance is found as an extra formant that amplifies the spectral level around the resonance frequency with no significant amplification on formants in the lower and higher frequency regions.

2.4.3.2 *The piriform fossae*

The piriform fossae are bilateral depressions, on each side of the larynx, formed in part by the aryepiglottis folds [66]. The pharynx terminates well below the epilaryngeal border, at the bottom of the piriform fossae (see *a tube within a tube*, 2.4.3.1). Painter [77] claims that the volume of the piriform fossae cannot be actively enlarged but action of the inferior pharyngeal constrictor muscles, posteroanterior expansion of the epilarynx, or raising the larynx can actively reduce their volume.

The piriform fossae, or piriform sinuses, owe their name to their pear shape. This pair of bilateral cavities is located posteriorly at the bottom of the pharynx, just above the oesophageal entrance [18].

Fant [32] made an observation about the role of the piriform fossae in the vowels formants. He used X-ray simulation to demonstrate that the spectral effect of the piriform fossae is to significantly lower the formants of vowels. Later, Sundberg [101] noted that the fossae play a significant role in the singer's formant between 2 and 3 kHz.

Baer et al. [5] estimated the vocal tract transfer functions with and without piriform fossae. They demonstrated an increase of the vowels

formants frequencies, due to the piriform fossae, in agreement with Fant's work. They suggested that their volume can be altered during articulation and become smaller during the production of back vowels. Davies et al. [21], using the data from Baer et al. [5], found a decrease of around 5% in F_1 , F_2 for the vowel /a/ when the fossae were incorporated as side branches to the vocal tract.

Dang and Honda [18] carried out a study of the piriform fossae on mechanical models as well as on human subjects, injecting water in the piriform sinuses of humans phonating in a supine position and in mechanical models of the lower half of Vocal Tracts. Comparing the acoustic output with and without piriform fossae they found that the fossae behave as side branches of the main tract and have a significant effect on the transfer function. They suggest that the piriform fossae need to be incorporated as side branches in any realistic model of speech production. For both models and humans, they found that the epilarynx tube resonance was enhanced, and that the fossae not only affected the spectral shape in the neighbourhood of its antiresonance but also decreased the lower resonance frequencies [18, 48].

The piriform fossae, as side branches of the Vocal Tract (VT), are thought to produce troughs in the region of 4 to 5 kHz [104, 48]. Titze and Story [108] found that the formant frequencies are slightly shifted when appending the piriform fossae to the main tract. In particular, they qualify the fossae as a *formant repellent*, generally pushing F_1 , F_2 , F_3 and F_4 lower and F_5 higher.

The piriform fossae usually show left-right asymmetry to various degrees [48]. Mokhtari et al. [68] suggested that when the bilateral fossae from two side branches with large asymmetry, vowel spectra result in two zero-pole pairs, while a single zero may be observed when the fossae are symmetrical.

SUM-UP A common consensus as to the acoustic effects of the hypopharyngeal cavities does not exist, as they have been related to affect F_4 and F_5 [32], to define F_3 and F_4 [101], to exert influence over

a wide frequency range above approximately 2.5 kHz [57], to produce a spectral peak at 3-3.5 kHz [48] or to define F₄ [104].

2.4.3.3 *The valleculae*

The valleculae epiglotticae are depressions (vallecula) just behind the root of the tongue; they constitute a borderline area between the pharynx and the larynx: these depressions serve as a temporary saliva reservoir to prevent from swallowing [10]. Physiologically, the valleculae play a role in feeding: they contribute to the process of swallowing by storing temporarily a bolus of food or liquid before it is propelled into the oesophagus [87].

Vampola et al. [115] reported for their female subject a large increase (64%) of the volume of the valleculae between before and after having phonated into a tube.

2.5 VOLUMETRIC IMAGING

In the last two decades, various imaging techniques have become available to characterise the vocal tract configurations of human subjects. Volumetric imaging, or 3D imaging, includes obtaining a contiguous set of image slices of the body part encompassing the vocal tract, segmenting the airway from its bordering tissues and subsequently reconstructing the vocal tract in three dimensions [114, 5, 91, 92]. Images can be acquired by Magnetic Resonance Imaging (MRI) or by Electron Beam Computed Tomography (EBCT) [93, 94, 115]. MRI has the clear advantage of presenting no adverse effects (no hazardous effects have been observed from short term exposure to magnetic fields used in MRI scanners [114, 91, 94, 100]). Due to the potential health hazards associated with exposure to ionizing radiations, it is no longer possible to use lateral X-Ray imaging (as in [32, 14]) for a nonmedical use [15], unless a consent form relating the potential hazards has been signed by the subject [115]. However, MRI has

some shortcomings when it is used to obtain images of airways [5], including limited image resolution and accuracy. Air-to-tissue boundaries can be distorted due to MRI artefacts [91], having the effect of blurring the edges of the vocal tract. Tissue with poor hydrogen density, such as bones and teeth also appear to be at the same gray scale density as air [91, 15]. Moreover, MRI acquisition time is long in comparison with EBCT (4-5 min [91] against 12-18 s [93]) [15]. On the other hand, EBCT has a higher resolution, a faster acquisition time, and allow teeth imaging, but necessitates the speaker/singer to be allowed only very low doses of radiation [94].

Several studies have used MRI to investigate the vocal tract shape during vowel phonation [91, 5, 92, 100], others used Computer Tomography (CT) [93, 94, 115]. Story et al [91, 92], Baer et al. [5] and Sulter et al. [100] used MRI to find the cross-sectional areas within planes perpendicular to the centreline extending from the glottis to the opening of the mouth to produce an *area function* (see 2.5.1).

More recently, Vampola et al. [115] used CT to assess the vocal tract shape changes due to exercising phonating into a tube. Echternach et al. [27] have used MRI volumetry to analyse the vocal tract configuration changes across registers (modal and falsetto voice) of 10 professional tenors. Laukkanen et al. [59] investigated the effects of phonation into a straw on the vocal tract adjustments and formant frequencies via MRI. Rua Ventura et al. [82] carried out an MRI-based study to assess the morphological differences in the Vocal Tract resonances of voice professionals.

A study conducted by Honda et al. [48] used refined MRI acquisition and visualisation techniques to highlight the significant effects of the hypopharyngeal cavities in speech.

Takemoto et al. [104] have developed a 3D cine-MRI technique to acquire the dynamic pattern of the vocal tract during the utterance of the short sequence of vowel /aiueo/ in Japanese, from which area functions were extracted frame-by-frame. A potential application of

this method is to measure the movements of the hypopharynx during changes of voice quality and fundamental frequency.

2.5.1 Vocal Tract Area

The Vocal Tract Area function is a useful representation of the VT from which the overall acoustical characteristics of the sound spectrum can be derived. Volumetric imaging (such as MRI or EBCT, see section 2.5) of humans phonating can be used to approximate the shape of the vocal tract by an *area function*: the variation of the cross-sectional area of the vocal tract as a function of the distance to the glottis [91, 92, 93, 94, 95, 97, 99]. The acoustic resonances of the tract can be calculated based on the shape of the area function, and will define the location of the formants [97]. Based on the assumption of one-dimensional wave propagation, a transmission line model [90] can be used, where the vocal tract is approximated by a series of cylinders whose dimensions are given by the area function. The cross-sectional areas are evaluated within planes perpendicular to the centreline which runs from the glottis to the mouth. The algorithm, called *iterative bisection algorithm*, works as follows [91]: a straight line is drawn between both extremities of the Vocal Tract (the glottis and the centroid of the cross-section made by the lips). The bisecting plane of this straight line delimits a cross-sectional area on the vocal tract. Taking the centroid of this cross-sectional area determines a new point. This delimits two segments to approximate the vocal tract. Iterating the process leads to both the creation of a centreline for the vocal tract and a set of oblique cross-sectional areas which will constitute the area function.

Story [97] uses the acoustic sensitivity function to allow the area functions to be modified so that their formant frequencies reach a set of targets. The sensitivity of a particular formant frequency to a

change in vocal tract cross-sectional area is defined as the difference between the kinetic energy (KE) and the potential energy (PE), as a function of the distance from the glottis, divided by the total energy of the system (TE) [33]. An acoustic sensitivity function can be written as

$$S_n(i) = \frac{KE_n(i) - PE_n(i)}{TE_n} \quad n = 1, 2, 3, \dots \quad \text{and} \quad i = [1, \dots, N_{\text{areas}}] \quad (2.3)$$

where i is the section number (being numbered such as section 1 is just above the glottis and section N_{areas} is at the lips), n the formant number, and

$$TE_n = \sum_{i=1}^{N_{\text{areas}}} [KE_n(i) + PE_n(i)] \quad (2.4)$$

The potential and kinetic energy for each formant are based on the volume velocity $U_n(i)$ and pressure $P_n(i)$ calculated for each section of the area function. They are computed as follows:

$$KE_n(i) = \frac{1}{2} \frac{\rho l(i)}{a(i)} |U_n(i)|^2 \quad (2.5)$$

and

$$PE_n(i) = \frac{1}{2} \frac{a(i)l(i)}{\rho c^2} |P_n(i)|^2 \quad (2.6)$$

where $a(i)$ and $l(i)$ are the cross-sectional area and length of element i , ρ and c are the density of air and the speed of sound respectively.

2.6 STUDYING THE VOCAL TRACT RESONANCES

This section reviews some techniques to measure Vocal Tract resonances. The ratio $p_m(f)/p_g(f)$ of the pressure at the mouth $p_m(f)$ to the pressure at the glottis $p_g(f)$ gives the transfer function of the tract $|T(f)|$, the relevant function to understand the timbre or the voice quality. Unfortunately, a direct measure would be too invasive

and therefore voice researchers have had to use different techniques which mostly fall into three categories: measure of the voice output itself, the use of an external stimulus, or an image-based model. A 4th category of VT resonances measurement has been used recently, the measurement of the resonances of an image-based 3D-printed VT.

2.6.1 Output sound after excitation at the glottis

A method consists of analysing the sound output when the VT is submitted to a glottal excitation such as in speech/singing. A common approach to estimate the VT resonances is to use the linear prediction of an autoregressive filter which would fit best the output spectrum [4, 63, 1]. The poles of this filter indicate approximately the resonances of the tract.

Quasi-periodic vibration of the VFs as in speech/singing is commonly used as a stimulus of the VT resonances. But it has two disadvantages: the source function, i. e. the glottal flow, is not known and the frequency resolution of the output depends on f_0 . For instance, for a fundamental frequency of $f_0 = 400$ Hz, the harmonics will be spaced by 400 Hz, not offering good frequency resolution.

Instead of using a harmonic spectrum as the source, it is possible to use a broadband signal, such as in whispering [79] or by ingressive phonation [65] for instance. This has the advantage of providing a continuous spectrum, whose frequency resolution is no longer dependent on the fundamental frequency of the source. The major drawback of this kind of method is that the glottal aperture in whispering and breathing in is usually bigger than during speech/singing. As a consequence, R_1 is risen [52]. Another inconvenience is that articulation may change from phonation in speech/singing to whispering, altering the resonances [39, 54].

The third method is to stimulate the VFs mechanically, from outside the neck, while the subject is phonating. Different signals have been used as excitation: sinusoids [40], white noise [12] or pseudo-random excitation [24]. One of the limitations is that the subject has to maintain the same articulatory position for a long time. Moreover, this method needs a high signal-to-noise ratio, in order to compensate for the damping of the signal when it travels through the skin and cartilage, which has been reported to be uncomfortable for the subject [24]. The transfer function from the outside of the neck to the larynx (skin, tissue, cartilage) is also unknown [79].

2.6.2 Output sound after excitation at the lips

Another method consists of measuring and analysing the response of the tract to an acoustical stimulation at the lips. Synthesised acoustic currents have been used for this purpose [28, 26, 53]. The method consists of injecting an acoustic current into the mouth, while simultaneously recording the sound pressure at the mouth. The ratio $Z = p_m(f)/U_m(f)$, gives the impedance at the mouth, with $p_m(f)$ the pressure measured at the mouth and $U_m(f)$ the acoustic current flow injected from the mouth. This method presents different disadvantages. First, the impedance measured at the mouth is in parallel with the radiation impedance. Therefore, the injected acoustic current needs not to be too weak, especially in the case of loud phonation at low frequency, because the radiation impedance is small for low frequencies. Second, it measures the vocal tract impedance, seen from the lips, and not from the glottis [119].

2.6.3 Numerical simulation of the transfer function of an image-based VT

The third approach is to use Magnetic Resonance Imaging or X-ray imaging to deduce the shape of the Vocal Tract and run acoustic models to obtain the transfer function [100, 5, 65, 92, 104]. One of the possible limitations is the many processes implied between the image acquisition and the production of a 3D model of the tract: image resolution and accuracy are limited [5], and air-to-tissue boundaries can be distorted due to MRI artefacts [91], having as effect to blur the edges of the vocal tract.

2.6.4 Measurement of an image-based 3D-printed VT

More recently, another method has been used to evaluate the resonances of the tract. It measures the transfer function from 3D printed replicas of MRI-based Vocal Tracts while phonating [106]. One of the limitations is the fact that the walls of the 3D vocal tract are rigid and have different absorption behaviour than real VTs [106].

2.6.5 Methods used in this thesis

The present thesis uses the last two methods: numerical simulation of the transfer function of an image-based VT and measurement of an image-based 3D-printed VT. This choice was made because image-based VT allows to use a refined representation of a real VT to investigate the spectral impact of its sub-structures on the acoustic output of the human voice. Furthermore, numerical and experimental methods are applied on an image-based VT: this provides benchmarking between both.

2.7 VOICE CLASSIFICATION

According to Titze [109], professional singers can usually be classified into six major categories, three per gender, i. e. low, medium and high voices [16]: from the lowest to the highest voice category, one finds Bass, Baritone, Tenor, Contralto, Mezzo-Soprano, Soprano. Traditionally, singing teachers tend to classify singers according to voice range, tessitura and timbre, which is a subjective classification method [109, 83, 84]. Titze suggested that objective measurements of the Vocal Tract dimensions including its length and volume could be significant to voice classification [109]. Voice researchers have been investigating factors such as f_0 , vocal fold length, formant frequencies, vocal tract dimensions and the Singer's Formant Cluster to sort singers into different voice types [8, 16, 25, 80]. [16] and [25] were the first to investigate the relationships between vocal tract dimensions, the SFC and voice types. [8] reported a correlation between the SFC and the voice type.

Cleveland [16] investigated the correlation between formant frequencies and voice classification of three vocal categories of male singers, i. e. bass, baritone and tenor. For this study, eight professional singers sang five different vowels, at the same pitch. The formant frequencies were directly measured from the sound output and singing teachers were asked to sort the eight singers into the three vocal categories. Results showed that the average four lower formants were correlated with the voice classification: voices with high formant frequencies were chosen as tenors, whereas voices with low formant frequencies were classified basses. Cleveland also found that formant frequencies in particular vowels gave a more significant voice classification than others.

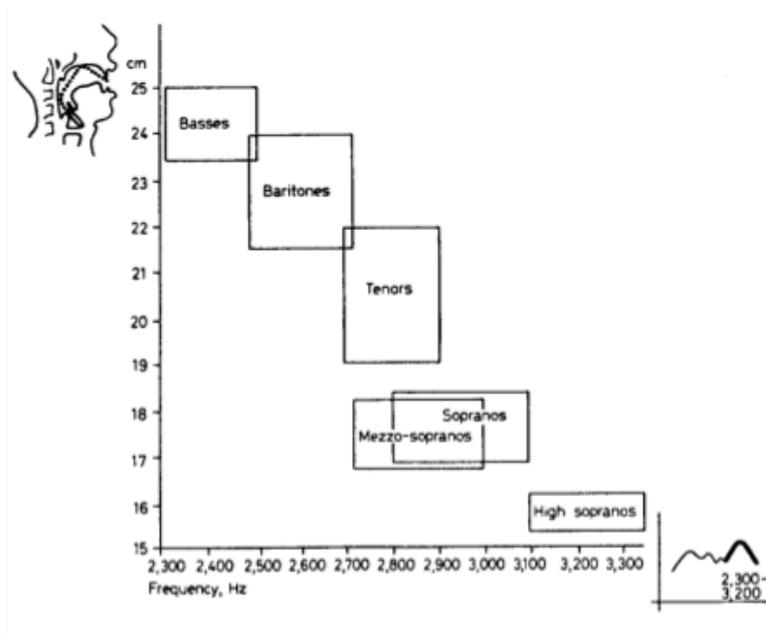


Figure 2.14: Relationship between the SFC and the dimensions of the VT, from [25]

Dmitriev and Kiselev [25] found that the dimensions of the VT were related to the production of low and high formants and to the voice classification. They measured the Vocal Tract length of 20 Russian professional opera singers from the Great Theatre with X-ray imaging. They also recorded the sound outputs to measure the formant frequencies. The results are shown in Fig 2.14 and Table 2.1: the shorter the VT, the higher the frequency of the high singing formant. From Basses, Baritones, Tenors, Mezzo-sopranos, Sopranos and high Sopranos are classified accordingly, with basses having the longest VT and lowest frequency of the high singing formant, up to the high sopranos with the shortest VT and highest frequency of the high singing formant. The second column of Table 2.1 gives a relation between the voice classification and the frequency of the high singing formant: the lower the voice classification, the lower frequency of the high singing formant.

Berndtson et al. [8] used a listening test with synthesised vowel stimuli to investigate the importance of the centre frequency of the

Voice type	Frequency of the low singing formant [Hz]	Frequency of the high singing formant [Hz]	Length of the Vocal Tract [cm]
High soprano	760-800	3100-3500	15.3-16.3
Soprano	700-760	2800-3100	16.8-18.5
Tenor	600-640	2700-2900	19.0-22.0
Baritone	540-600	2500-2700	21.5-24.0

Table 2.1: Frequency of the low singing formant, frequency of the high singing formant, and the length of the vocal tract for different voice categories, from [25]

singer's formant to voice classification. They found that the centre frequency of the singer's formant clearly influenced the categorisation of synthesised voices into bass, baritone, tenor, and contralto voice categories. They also reported that the centre frequency of the singer's formant was relevant to the naturalness of soprano synthesis.

Roers et al. [80] studied the relationship between the VF length and voice classification. They measured the larynx and VF length of 132 professional singers with X-rays, of which there were 40 sopranos, 22 mezzo-sopranos, 9 altos, 19 tenors, 23 baritones and 19 basses. They also measured directly the VF length of 29 singers. The results showed that VF length is strongly correlated with the anterior-posterior diameter of the subglottis and trachea. The data was used to predict the VF length of the 132 singers and a covariation between the VF length and voice classification was found, yielding to an average of 14.9, 16.0, 16.6, 18.4, 19.5, and 20.9 mm for sopranos, mezzo-sopranos, altos, tenors, baritones and basses respectively. The data support the fact that there are consistent anatomical differences amongst singers of different voice categories.

In a recent study, Yan et al. [120] have used an acoustic pharyngometer, a non-invasive device, relying on Acoustic Reflection Technology (ART) which can directly measure the vocal tract dimensions and volume. It is very similar to a sonar in that it sends a sound wave into the airway. A fraction of the acoustic wave is reflected back at each point of discontinuity in the upper airway and is recorded by a microphone [56]. The signal can then be transformed into dimensional data. The changes in cross-sectional area can then be used to deduce the length and volume of (a portion of) the tract. Yan et al. [120] suggest that the voice classification is related to the length and volume of the vocal tract: higher voices exhibit a smaller VT volume and higher formant frequencies than lower voices.

Part II

MATERIALS AND METHODS

To analyse the spectral effect of the hypopharynx on the singing voice, the Vocal Tract (VT) will be approximated successively by four types of Vocal Tract Models (VTMs) as can be seen on Fig 2.15: in order of increasing complexity/realism, VTM-1, VTM-2, VTMCh&K, VTM-MRI:

1. **VTM-1**: a single tube, of length 142 mm, radius 30 mm and thickness 2 mm, representing the Vocal Tract proper (VTp), defined as the VT without the epilarynx (introduced by [104], see section i). Note that the VTp is termed oropharynx in the section iii of this thesis.
2. **VTM-2**: a twin-tube, composed of VTM-1 and an appended tube representing the epilarynx, whose length and radius are changed from 16 to 28 mm and from 2 to 9 mm respectively.
3. **VTM-Ch&K**: variable cross-sectional tube, from the average cross-sectional area of the Vocal Tract derived from X-Rays of Japanese speakers [14].
4. **VTM-MRI**: actual VT shape acquired by MRI of six professional singers.

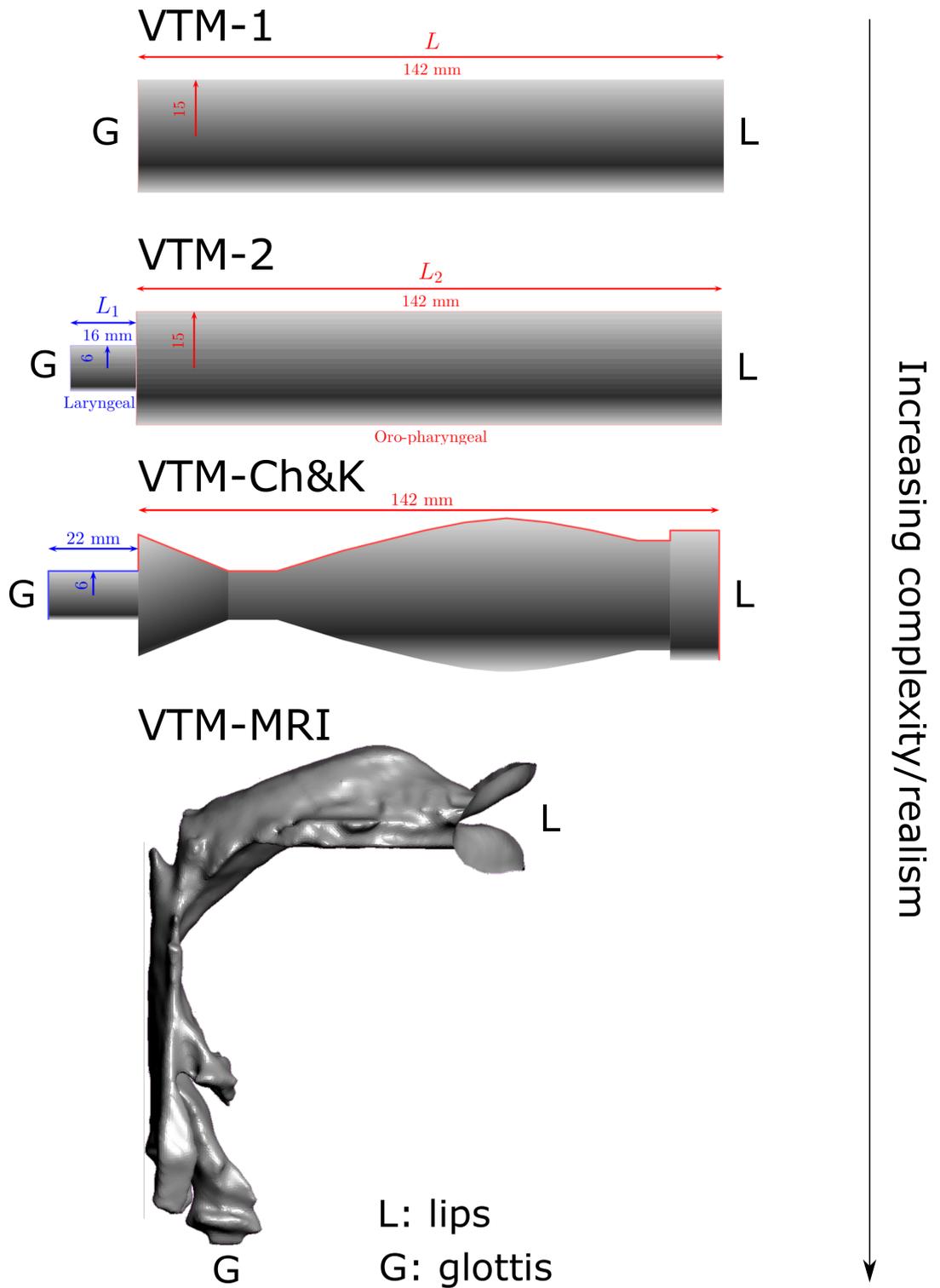


Figure 2.15: Vocal Tract Models used in this thesis. By order of increasing complexity/realism, VTM-1, VTM-2, VTM-Ch&K and VTM-MRI.

NOTE1 : The VTM-Ch&K represent the 5 Japanese vowels /a/, /e/, /i/, /o/ and /u/.

NOTE2 : The VTM-MRI are based on the VTs of 6 professional singers (spread across voice categories). The corpus is composed of 1 Soprano, 2 Mezzo-Sopranos, 1 Tenor, 1 Bari-Tenor and 1 Bass-Baritone. In order to maintain anonymity, each singer has been assigned a name which mnemonically serves to indicate their voice type.

- **BarnaBy** is a **Bass-Baritone**, aged 31.
- **BarTholomew** is a **Bari-Tenor**, aged 34.
- **Timothy** is a **Tenor**, aged 30.
- **MariStela** is a **Mezzo-Soprano**, aged 29.
- **MariSa** is a **Mezzo-Soprano**, aged 35.
- **Sophy** is a **Soprano**, aged 30.

The data of Barnaby, Bartholomew and Maristela were acquired in the York Neuroimaging Centre (YNiC) of the University of York, in the United Kingdom (labelled protocol "York" in this thesis). The other set of MRI data (Sophy, Marisa and Timothy) were acquired in the Department of Radiology of the University Medical Center Freiburg, in Germany (labelled protocol "Freiburg" in this thesis). Both protocols are detailed in subsection 5.1.1, page 103. For more details about the singers, refer to Table 2.2, which shows their age, classification according to the Bunch and Chapman criteria [11], range, sung pitch and vowel and protocol acquisition.

NOTE3 : Bartholomew sang on 5 different registers named according to his own terminology: **Chest**, on F#3; **Mixed**, on F4; **Squillo**, on A4; **Rinforzando**, on B4; **Falsetto**, on A4.

Singers data

	Age	Classification	Range	Pitch	Vowel	Protocol
Sophy	30	National	G3-D6	G5	/hard/	Freiburg
Maristela	29	National	F3-C#5	C4	/hard/	York
Marisa	35	International	D3-A5	A4	/hard/	Freiburg
Timothy	30	National	F2-C#5 (E5)	F4	/hard/	Freiburg
Bartholomew	34	International	E2-D5 (G5)	F#3	/hard/	York
Barnaby	31	National	C2-A4 (A5)	G#2	/hard/ /port/ /stern/ /food/ /neap/	York

Table 2.2: Age, classification according to the Bunch and Chapman criteria [11], range, sung pitch (between brackets for falsetto), vowel and protocol acquisition for the 6 professional singers.

NOTE 4 : To benchmark the numerical method, some VTM-MRI were 3D-printed. Only a few could be printed due to budget restrictions. According to [85], the data of Barnaby were acquired during a rather stable phonation which defines clear edges between structures on the MRI images. Therefore, the VTM-MRI of Barnaby was chosen for printing. The five vowels printed are those as in /hard/, /port/, /stern/, /food/ and /neap/.

The different VTMs used for this thesis can be visualised on the diagram of Fig 2.16.

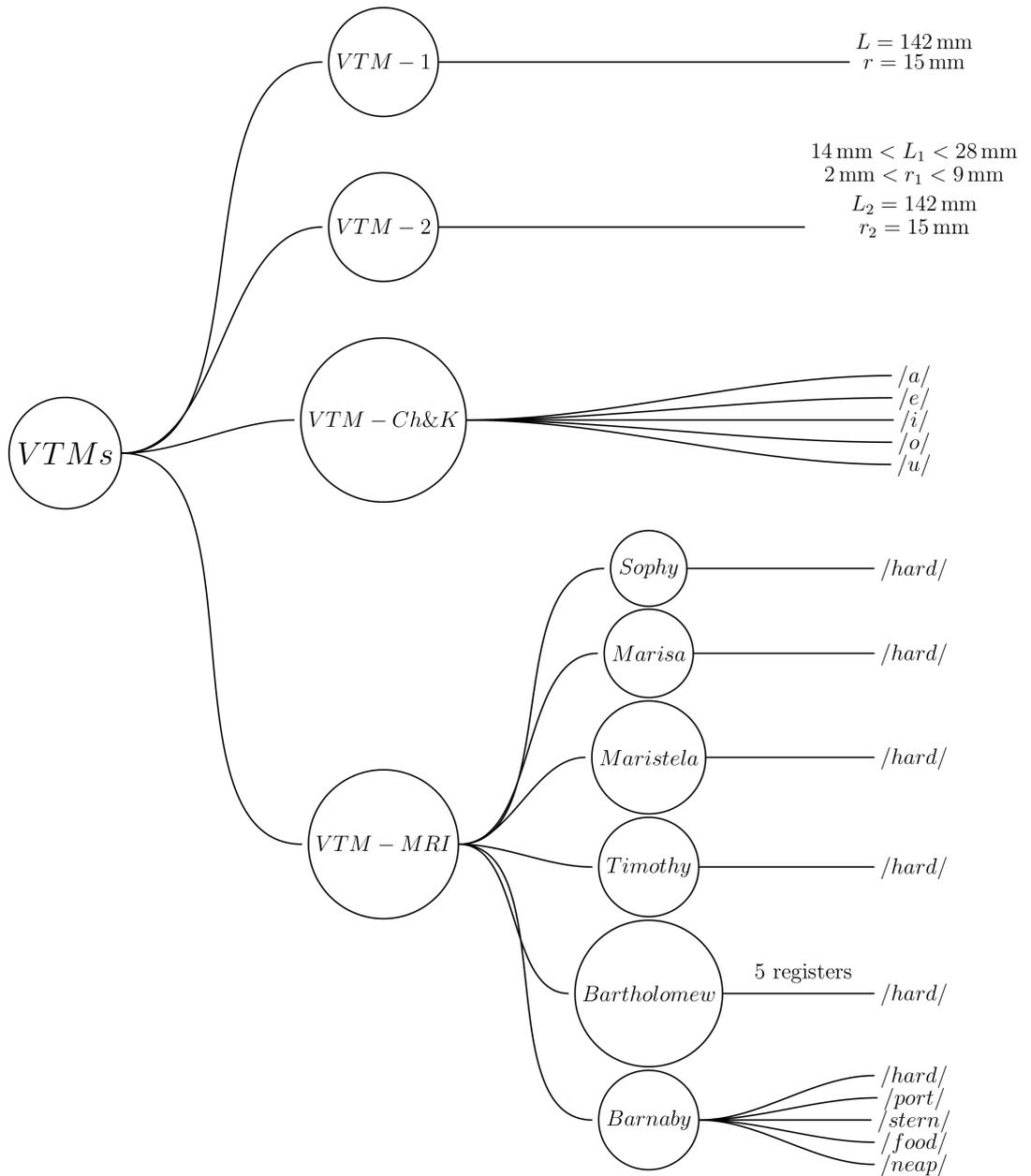


Figure 2.16: Diagram of the different VTMs used in the thesis. By order of complexity/realism, VTM-1 (1 tube), VTM-2 (2 tubes), VTM-Ch&K (based on Chiba & Kajiyama's [14]) and VTM-MRI, actual VT shape acquired by MRI of six professional singers. Note that Bartholomew's VT has been processed for 5 registers and Barnaby's VT for 5 vowels.

This section is divided into three parts:

- Numerical simulations
- Experimental measurements
- From MRI to 3D printing

The first section refers to the numerical methods chosen, i. e. the Finite Volume Method (FVM) which works in the time-domain and the Finite Element Method (FEM), which works in the frequency-domain. The open-source code OpenFOAM (see section 3.2, page 64) implements the FVM whereas the software ACTRAN is used for FEM.

The second section develops a novel method to experimentally measure the impulse response of a cavity and derive its transfer function. This approach is inspired from the method used by Farina [34] in room acoustics.

The third section lists and develops the steps from the acquisition of MRI data of professional singers phonating in a supine position to the segmentation, meshing and 3D-printing of their reconstructed Vocal Tract.

3 | NUMERICAL SIMULATIONS

This chapter explains the numerical methods used to compute the transfer functions of the different VTMs. The **Finite Volume Method** (FVM) and **Finite Element Method** (FEM) are both largely used in a variety of fields, such as automotive engineering, aerospace, aircrafts, civil engineering, military engineering, architecture, etc [29]. It first introduces the notion of computational domain, then describes the two methods used: **FVM** and **FEM**. For extensive development of these two methods, refer to Appendix B and C respectively. The chapter closes with a section about meshing; an indispensable tool to turn real life objects into usable geometric data for FVM/FEM.

3.1 COMPUTATIONAL DOMAIN

For both **FVM** and **FEM**, the computational domain or domain of computation represents the physical space on which the phenomenon of interest (here, the lossless wave propagation through a cavity closed at one end and opened at the other) is simulated. This domain is enclosed by boundaries sorted in different boundary condition patches (see Boundary Conditions in the next sections). For the present case, the walls and glottis end of the VTMs are set as walls, offering a reflection of the waves, according to a preset absorption coefficient, and the radiating end (as an extension of the lips end) exhibits a reflection-free boundary. For an example of the computational domain for VTM-1, see Fig. 3.1.

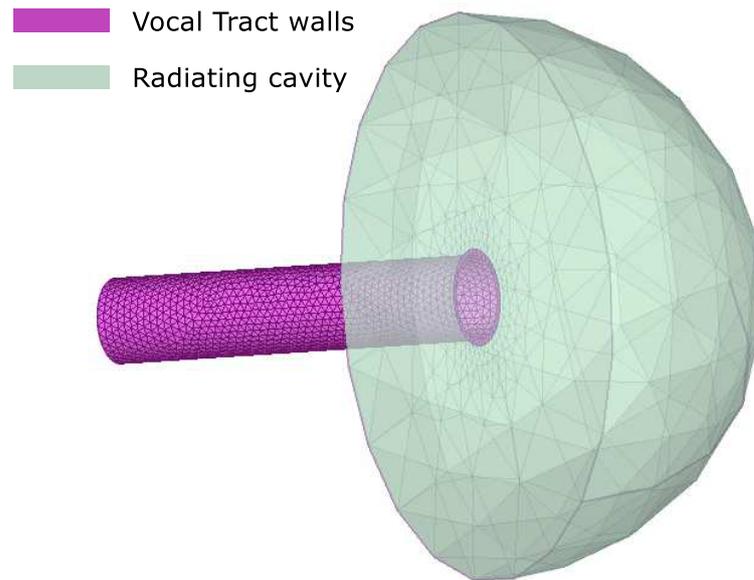


Figure 3.1: Computational domain of VTM-1: separation of the boundary between two patches: the VT walls and the radiating cavity.

3.2 FINITE VOLUME METHOD

The open source code OpenFOAM® (Open Field Operation and Manipulation, www.openfoam.com) was used to implement the lossless wave propagation of an impulse inside the VTMs. The code being open and modifiable, OpenFOAM offers users complete freedom to customise and extend its existing functionality. For the purpose of this thesis, a lossless wave equation was implemented on OpenFOAM, along with suitable boundary conditions. The numerical schemes used (see hereunder, Euler Explicit and Gauss Linear) were the ones included in the software. OpenFOAM is a Computational Fluid Dynamic (CFD) software working in the time domain with the FVM. For more extensive development about the Finite Volume Method (FVM), refer to Appendix B.

3.2.1 Wave equation

The 3D (incompressible) wave equation is as follows:

$$\frac{\partial^2 \phi}{\partial t^2} = c^2 \nabla^2 \phi \quad (3.1)$$

where ϕ is a scalar potential, t is the time and c is the speed of sound. If p is the acoustic pressure and u the acoustic particle velocity, ϕ can be defined such as:

$$\begin{aligned} p &= \frac{1}{\rho} \frac{\partial \phi}{\partial t} \\ u &= -\nabla \phi \end{aligned}$$

where ρ is the medium density, so that (3.1) is considered to represent a linear sound propagation [81]. (3.1) is implemented in OpenFOAM on the whole domain, and solved with Euler explicit (for more details about the numerical scheme, refer to Appendix B) for the time derivative and Gauss linear for the Laplacian.

3.2.2 Impulse Response

The FVM works in the time-domain. Therefore, the impulse response is first numerically simulated and then transformed into the frequency-domain in order to obtain the transfer function. The initial and boundary conditions are set as follows:

3.2.2.1 Initial condition

At the time $t = 0$, the velocity is equal to zero everywhere and the pressure is equal to the atmospheric pressure everywhere except at the glottis with a point source, implemented under the form $e^{-\|(x,y,z)\|}$, to elicit the propagation of an impulse.

3.2.2.2 Boundary Conditions

The domain is divided into two different boundary conditions, i.e. the Vocal Tract walls (including the closed glottis) and the radiating cavity.

This thesis uses the so-called "convective outlet", an incompressible reflection-free boundary condition taken from [107].

The convective outlet used for the wave equation boundary condition is given by

$$\frac{\partial \phi}{\partial t} + V_n \frac{\partial \phi}{\partial n} = 0 \quad (3.2)$$

with the convective velocity V_n represented by

$$V_n = \frac{Z_n}{\rho} \quad (3.3)$$

and Z_n being the normal acoustic impedance relating to the absorption factor α by

$$\alpha = 1 - \left| \frac{(Z_n/\rho c) - 1}{(Z_n/\rho c) + 1} \right|^2 \quad (3.4)$$

NOTE

- The Vocal Tract walls are assigned an absorption coefficient of 0.01[76].
- The open end, which radiates the sound outwards, is assigned an absorption coefficient of 1. This allows for a reflection-free boundary at the end of the radiating field [76].

The boundary conditions are as follows:

1. Vocal Tract walls

$$\frac{\partial \phi}{\partial t} + 136791.7 \frac{\partial \phi}{\partial n} = 0 \quad (3.5)$$

2. Radiating cavity

$$\frac{\partial \phi}{\partial t} + 343.7 \frac{\partial \phi}{\partial n} = 0 \quad (3.6)$$

The values in 3.5 and 3.6 have been computed for a speed of sound of $343.7 \text{ m} \cdot \text{s}^{-1}$ and a density of $1.205 \text{ kg} \cdot \text{m}^{-3}$ [81].

3.2.3 Transfer function

The impulse response (pressure recorded by a virtual microphone after 0.04 s, which was the minimum time duration to obtain a good resolution spectrum) was converted into a transfer function with an FFT, with no windowing. Note that the virtual microphone is placed 3 cm far from the centre of the cross-sectional area of the VT at the lips end (see Fig 3.2). This matches the condition set further in the experimental measurements, in section 4.

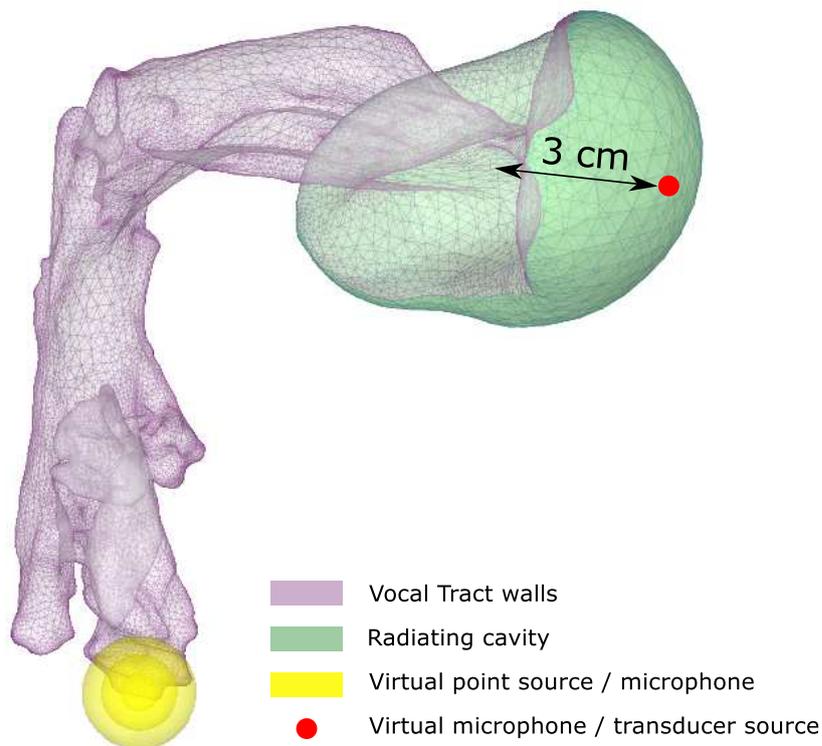


Figure 3.2: Computational domain for VTM-MRI-Barnaby-/stern/, with the different boundary conditions (Vocal Tract walls and Radiating cavity). The virtual probe microphone is located 3 cm far from the centre of the cross-sectional area of the lips opening and the point source (in yellow) is located at the glottis. Note that in the case of the experimental measurement of the 3D printed VTM-MRI, the (transducer) source and the microphone are interchanged.

3.3 FINITE ELEMENT METHOD

The software ACTRAN by FFT (Free Field Technologies, www.fft.be) was used to run the FEM simulations. This commercial CFD code is based on FEM. The user can simulate various cases of acoustic propagation by entering different parameters for the boundaries, excitation, material, etc. For an extensive development about the FEM, refer to Appendix C.

3.3.1 Transfer function

Actran computes the frequency response of the system to a given excitation (here, a source point at the glottis) for a range of frequencies. The equation solved by Actran at each frequency is:

$$(K + i\omega C - \omega^2 M) \mathbf{x}(\omega) = F(\omega) \quad (3.7)$$

with the stiffness matrix K , the admittance matrix C and the mass matrix M . Actran solved this equation for the range 20-10000 Hz by steps of 20 Hz. The response of the system to each frequency is computed by a virtual microphone situated 3cm far from the centre of the cross-sectional area of the VT at the lips (see Fig 3.2). The transfer function then simply consists of the adjunction of each frequency response of the system.

3.3.1.1 Boundary conditions

The absorption coefficients of the different boundaries are set respectively to:

- $\alpha = 0.01$ for the walls of the tract.
- $\alpha = 1$ for the open end, which radiates the sound outwards.

The values are those prescribed by Oshima et al. [76]. Note that the absorption coefficient of the walls has been adjusted to the value

of $\alpha = 0.02$ for the MRI-based Vocal Tracts, to match the experimental measured bandwidth of the formants, referring to the method used in [71].

3.4 MESHING

In the Appendices B and C, the physical domain Ω needs to be divided into small cells, Ω_i , $i = 1, \dots, N$ with $\Omega = \bigcup_i \Omega_i$. This operation is called *meshing*. In this research, the open source Salome is used for the meshing.

SALOME (<http://www.salome-platform.org>) is an open-source software that provides a generic platform for Pre- and Post-Processing for numerical simulation.

Scripts in Python were written to build the different VTMs in the geometry module of Salome and to generate meshes in the mesh module with NetGen 1D-2D-3D algorithm for tetrahedralisation and triangulation, with manual settings in the zones of interest. The VTM-MRIs are directly imported into the mesh module of Salome and meshed with NetGen automatic tetrahedralisation.

3.5 BLOCK-DIAGRAM

The different processes at work can be seen on the block-diagram 3.3. The diagram is divided into 3 main processes:

- Pre-processing
 - **Salome** is provided with python scripts which create geometric models, meshes and boundary condition patches (*VTM-1*, *VTM-2* and *VTM-Ch&K*).

- **ITK-Snap** (see section 5 segments MRI data into a volume and exports it to STL (*stereolithography* format) (*VTM-MRI*).
 - **Blender** inverts the face orientation of the STL file (*VTM-MRI*).
 - **NetGen** smooths the surface of the STL with automatic triangulation (*VTM-MRI*).
 - **Blender** is used to manually select the faces of the STL which correspond to the boundary patches (*VTM-MRI*).
 - **Salome** is used to mesh the corresponding STL, assigning boundary conditions to each patch (*VTM-MRI*).
- Simulation
 - **OpenFOAM** is implemented with a lossless wave equation, boundary conditions and initial condition on the computational domain defined by the mesh and boundary patches provided by Salome (*VTM-1*, *VTM-2*, *VTM-Ch%K*).
 - **ACTRAN** is given boundary conditions and initial condition on the computational domain defined by the mesh and boundary patches provided by Salome (*VTM-1*, *VTM-2*, *VTM-Ch&K*, *VTM-MRI*).
- 3D printing
 - **Salome** provides geometric models under the .STL format to be printed (*VTM-1*, *VTM-2*, *VTM-Ch%K*).
 - **ITK-Snap** provides .STL file from the segmentation of MRI data (*VTM-MRI*).
 - **Magics** edits the STL file (see section 5) (*VTM-MRI*):
 1. Creates a shell (give the STL surface a thickness)
 2. Splits the VTM in two parts, to make it foldable, for research purposes.
 3. Creates a hinge, to deal with the opening of the VTM.

4. Appends an "insert" so that it can be plugged onto a cabinet, for demonstration purposes.

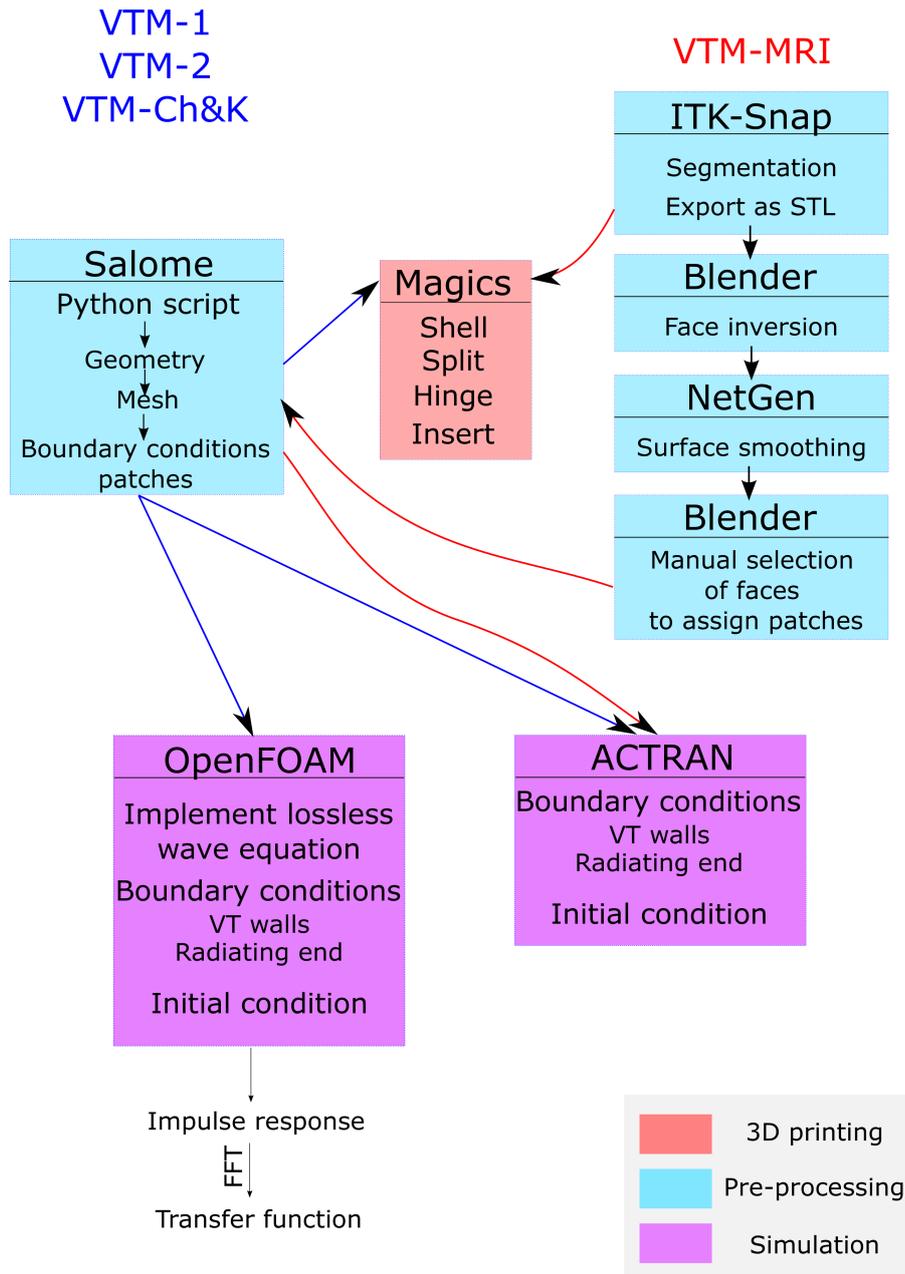


Figure 3.3: Block-diagram of the processes at work for the numerical simulations and 3D printing of VTM-1, VTM-2, VTM-Ch&K and VTM-MRI. The processes are divided into three categories: the pre-processing, the simulation and the 3D printing. For VTM-1, VTM-2 and VTM-Ch&K (blue arrows): a script in Python is written in Salome to create the different geometries and subsequent meshes, assigning boundary conditions to the patches which are then used both for FVM (OpenFOAM) or FEM (Actran), which implement the equation to be solved as well as the boundary and initial conditions on the computational domain. VTM-MRI follows the red arrows: ITK-Snap is first used to segment the MRI data from the singers, exported into Blender which inverts the face orientation, then to NetGen which automatically meshes the volume with tetrahedrons, then to Blender, which manually selects the faces to split the domain into patches which will be assigned different boundary conditions in Salome.

4 | EXPERIMENTAL MEASUREMENTS

This chapter is about the experimental set-up used to measure the VT transfer functions, as described in Fig 4.1. First, the method used to measure the transfer function of a cavity, including the input signal and its inverse filter, will be explained step by step. This will be followed by further explanation regarding the material used for the measurement, i.e. the speaker, the cabinet and the microphone.

4.1 METHOD

The method used to measure the impulse response and consequently the transfer function of the different VT models (VTM-1, VTM-2, VTM-Ch&K and VTM-MRI) is based on the methodology developed by Farina in [34] to measure simultaneously the linear impulse response and harmonic distortions of a room with an exponential sine sweep. This method is applied to a cavity, the VTM (Vocal Tract Model).

Fig 4.1 is an overview of the method which will be developed in further detail in the following sections:

1. A driver is given an input signal, ESS , which is recorded through a probe microphone (see section 4.2 page 76).
2. The output recorded by the microphone is then convolved with the inverse filter of the input signal, i.e. ESS^{-1} .

3. As a result, the impulse response is "linearised", i.e. the Linear Impulse Response (LIR) and the harmonic distortions are split apart.
 4. An FFT is performed on the LIR, giving the pressure-pressure transfer function.
 5. The transfer function of the driver alone is subtracted from the one with the VTM, giving as a final result the transfer function of the VTM, independent of the driver's frequency response.
- NB: in this thesis, processes 2 to 3 are termed "Linearisation of the impulse response"

The processes (1 to 4 in Fig 4.1) are repeated twice: once with the VTM, once without. The spectra obtained are then subtracted from one another (5 in Fig 4.1) to provide the transfer function of the VTM. The driver is located 3 cm from the centre of the cross-section of the lips opening (driver+VTM) or from the microphone standalone to match the condition of the numerical simulations of section 3 (see Fig 3.2).

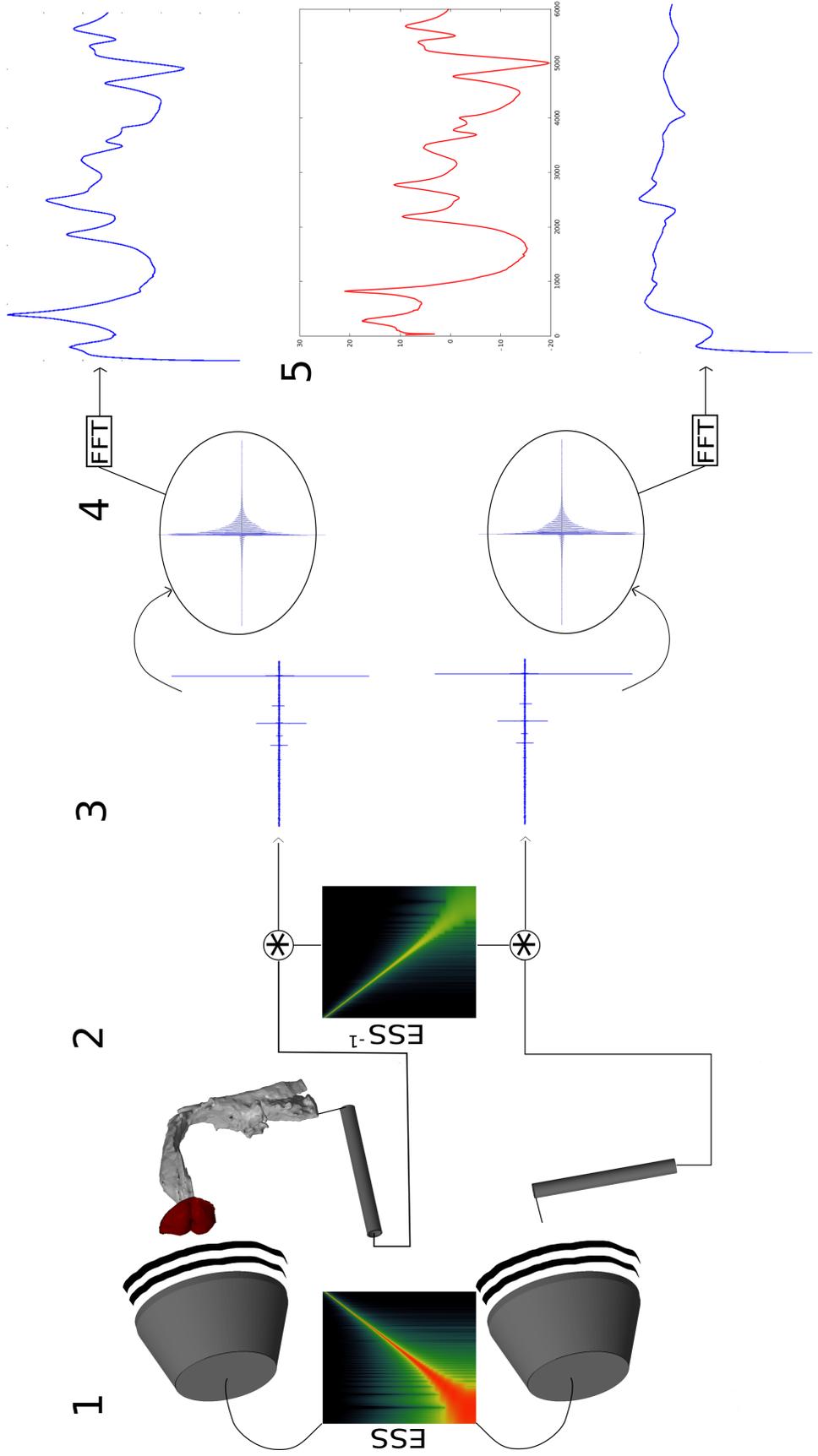


Figure 4.1: An Exponential Sine Sweep (ESS) is given as an input signal to the driver (1). The output recorded via a probe microphone is convolved with the inverse filter (ESS^{-1}) (2). It results in a temporal separation of the Linear Impulse Response (LIR) and the harmonic distortions (3). An FFT is performed on the LIR to give the linear transfer function of the system (4). Processes 1 to 4 are repeated twice: once with the Vocal Tract, and once without. Both spectra are then subtracted (logarithmic vertical scale) to give the transfer function of the standalone Vocal Tract.



Figure 4.2: 6-sided anechoic chamber.

4.2 MATERIAL

The experiment was carried out in a 6-sided anechoic chamber (see Fig 4.2). A probe microphone G.R.A.S. type 40SA was located at the glottis end. The signal was pre-amplified by a power Module type 12AA before being written on a USB type device with a 96 kHz sampling rate on a 24 bits WAV file. A transducer is situated 3 cm far from the radiating end (lips end) of the VTMs and radiates the sine sweep towards the VTM. This matches the condition set in section 3. The sine sweep is 10 s long and is swept from 10 Hz to 11000 Hz.

4.2.1 The driver

To measure the transfer functions of the different VTMs, a driver exhibiting a flat frequency response ($\pm 3\text{dB}$ between 0.1-10 kHz) was chosen, as can be seen in Fig 4.3. Then, a cabinet was designed to enclose the driver. The cabinet can be seen in Fig 4.4.

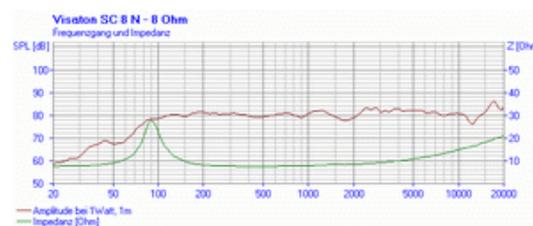


Figure 4.3: Frequency response of the driver VISATON SC 8 N 8 Ω

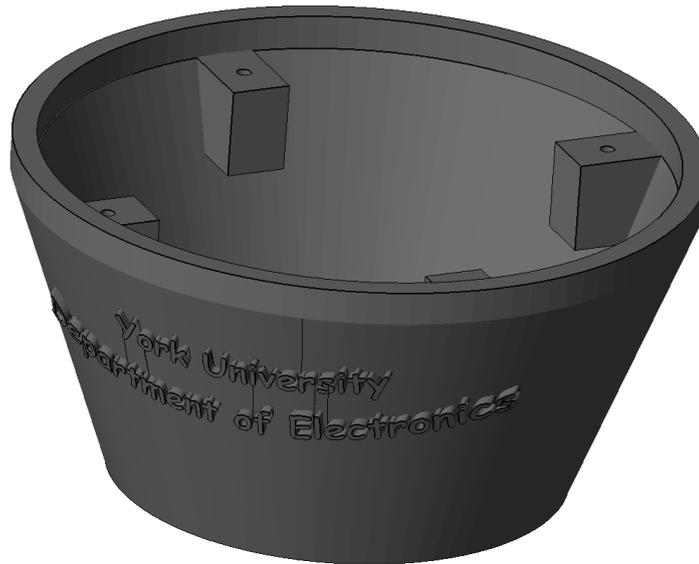


Figure 4.4: 3D-printed cabinet, designed to enclose the VISATON driver

4.3 EXPONENTIAL SINE SWEEP

Audio measurements, whether for room acoustics or audio equipment, are represented by two kinds of measurements: the linear transfer function of a system and its harmonic distortions [34]. Amongst the well-known methods, it is found in the literature that periodic pulse or Maximum Length Sequence (MLS) have been widely used as excitation signals; using a periodic pulse has proved to have a poor SNR (Signal-to-Noise Ratio) since the stimuli need to have a low energy to remain within a linear frame, i.e. to prevent from non-linear distortions[64]. MLS is usually employed to improve SNR, but has been shown to be inefficient when the non-linearities are too important [34]. These traditional methods prevent the separation of the linear impulse response from the harmonic distortions [64].

Another family of signals used for acoustic measurements, the swept sines, vary the frequency continuously and allow a better SNR and provide ways to separate the linear impulse response from its

harmonic distortions by a linear convolution of the output signal with the analytical inverse filter processed from the excitation signal [44]. The most common sweeps can be performed in two ways: linearly or exponentially [70].

The linear sweep increases its frequency with a fixed rate:

$$\frac{\omega_2 - \omega_1}{t_2 - t_1} = \text{const}$$

and it has a *white spectrum*: it contains the same amount of energy at each frequency.

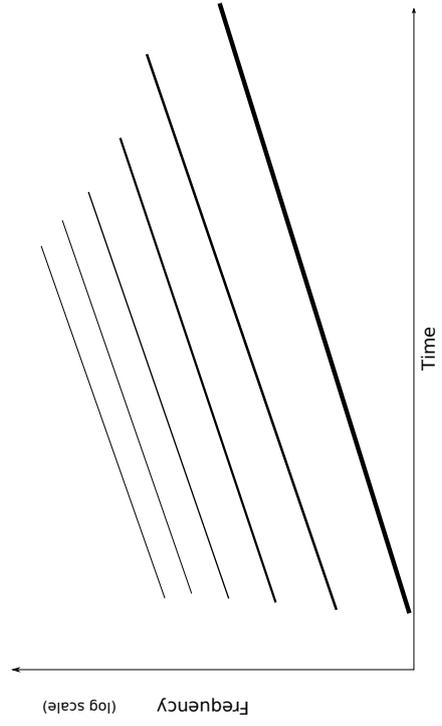
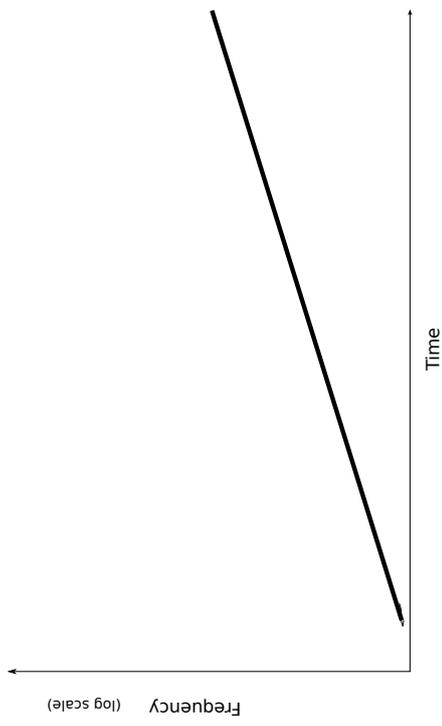
The exponential sweep increases its frequency with a fixed rate of the ratio of two frequencies:

$$\frac{\ln(\omega_2/\omega_1)}{t_2 - t_1} = \text{const}$$

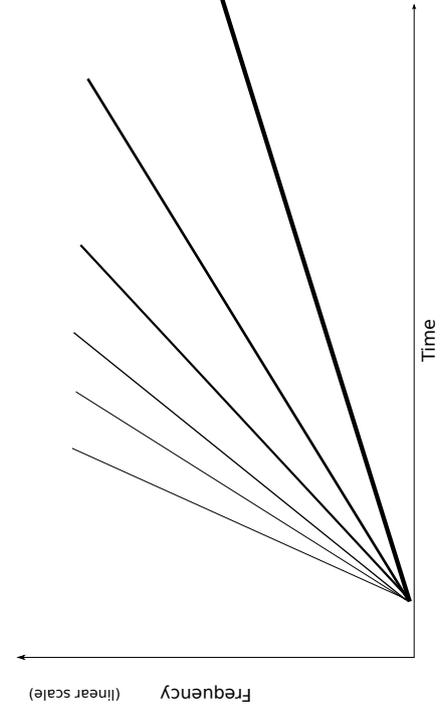
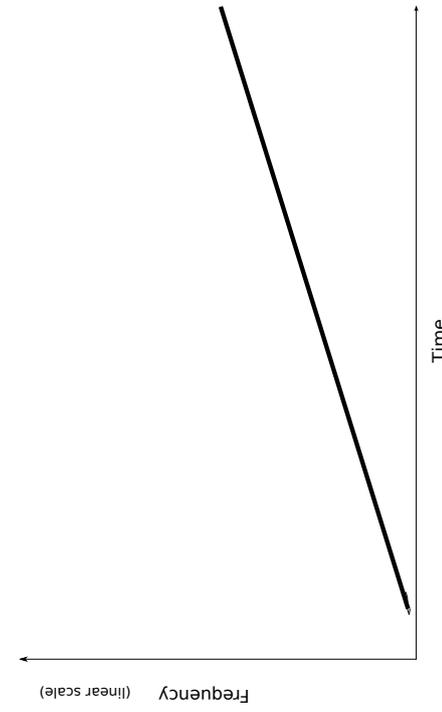
and it has a *pink spectrum*: it contains the same amount of energy per octave.

The linear sweep spends the same time between 20 Hz and 40 Hz and between 100 Hz and 120 Hz whereas the exponential sweep spends the same time between 20 Hz and 40 Hz (an octave) and between 10 kHz and 20 kHz (also an octave). This results in a white (flat) spectrum for the linear sweep, i.e. every frequency receives the same energy, and a pink (not flat) spectrum for the exponential sweep, i.e. the energy per octave is constant. The main advantage of the exponential sweep over the linear sweep is its possibility to separate the linear response from its harmonic distortions, as shown in Figs 4.5a and 4.5b (representation of spectrograms adapted from [34]): the system response to an exponential sweep excitation is composed a linear response and its harmonic distortions. Since the harmonic distortions are parallel to the linear response, both can be split apart by convolution. This is not possible for the linear response for which the linear response and harmonic distortions are not parallel. Fig 4.5 represents

an actual spectrogram of the system response to an exponential sine sweep. The most powerful (the amplitude is represented by a colour scale) is the linear response, the line on the bottom right.



(b) Representation of an exponential sine Sweep updated from Farina [34]: excitation signal (above) and system response (below) in the case of a weakly nonlinear system, displaying harmonic distortions.



(a) Representation of a linear sine Sweep updated from Farina [34]: excitation signal (above) and system response (below) in the case of a weakly nonlinear system, displaying harmonic distortions.

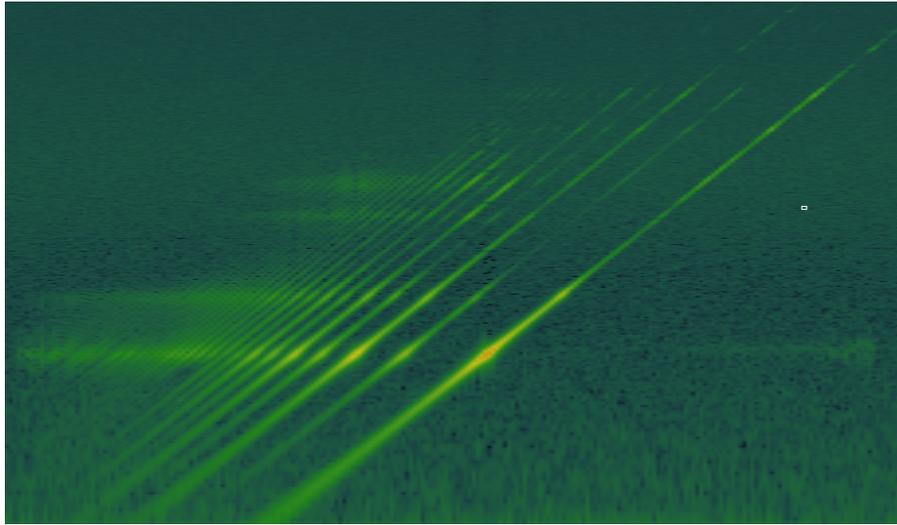


Figure 4.5: Spectrogram of the system response to an exponential sine Sweep in the case of a weakly nonlinear system, displaying the linear response of the system (main line on the bottom right) and its harmonic distortions (parallel lines to the main line). The amplitude is represented by a colour scale.

Generally, an exponential sine sweep is written in the form:

$$s(t) = \sin [\theta(t)] = \sin \left[K \cdot \left(e^{\frac{t}{L}} - 1 \right) \right] \quad (4.1)$$

Posing

$$\left. \frac{d \left[K \cdot \left(e^{\frac{t}{L}} - 1 \right) \right]}{dt} \right|_{t=0} = \omega_1 \quad \left. \frac{d \left[K \cdot \left(e^{\frac{t}{L}} - 1 \right) \right]}{dt} \right|_{t=T} = \omega_2$$

determines

$$K = \frac{T \cdot \omega_1}{\ln \left(\frac{\omega_2}{\omega_1} \right)}, \quad L = \frac{T}{\ln \left(\frac{\omega_2}{\omega_1} \right)}$$

where ω_1 and ω_2 represent the lower and upper frequency of the sweep respectively. T is the time duration of the sweep, in seconds, and t the time. The instantaneous frequency $\omega(t)$ (the first time derivative of $\theta(t)$) is then given by

$$\omega(t) = \frac{d [\theta(t)]}{dt} = \frac{K}{L} \cdot e^{\frac{t}{L}} \quad (4.2)$$

Let us now examine the time delay Δt needed by the sweep to go from one frequency f to N times this frequency, i.e.

$$\omega(t + \Delta t) = N\omega(t) \quad (4.3)$$

Expressing (4.3) according to (4.2), we find that

$$\frac{K}{L} \cdot e^{\frac{t + \Delta t}{L}} = N \cdot \frac{K}{L} \cdot e^{\frac{t}{L}}$$

from which we find

$$\frac{\Delta t}{L} = \ln N$$

and finally

$$\Delta t = L \cdot \ln N \quad (4.4)$$

which means that Δt remains constant once the order N has been fixed. In other words, the time Δt needed by the sweep between a pitch of frequency f and the same pitch 3 octaves higher, i.e. $2^3 \cdot f = 8 \cdot f$, Δt will be given as $\Delta t = L \cdot \ln 8$ whatever the frequency f is. Indeed, as described before, the sweep spends the same time between 20 Hz and 40 Hz (one octave) and between 10 kHz and 20 kHz (also one octave).

For a frequency-varying signal, the energy at a specific frequency is proportional to the time duration during which the signal oscillates at that specific frequency [64]. Since the energy signal $E(t)$ is proportional to the time at a specific frequency, it follows that $E(t)$ is inversely proportional to the rate of change of the instantaneous frequency $\omega(t)$. Thus, it follows that

$$E(t) \propto \frac{1}{\frac{d[\omega(t)]}{dt}} = \frac{L^2}{K} \cdot e^{-\frac{t}{L}} \quad (4.5)$$

Taking the Fourier transform of (4.5) leads to a formulation of the energy as a function of frequency:

$$E(j\omega) \propto \frac{L^2}{K} \cdot \frac{1}{L + j\omega}$$

Defining k as a constant of proportionality, the energy can be expressed as:

$$E(j\omega) = \frac{kL^2}{K} \cdot \frac{1}{L + j\omega} \quad (4.6)$$

This formula illustrates the important fact that the energy of an ESS decreases with frequency. This can be physically explained by the fact that the sweep spends more time at low frequencies than at high frequencies (cf. the example above with 20-40 Hz and 10-20 kHz). In particular, if we double the frequency, i.e. if we replace the factor $\frac{1}{\omega}$ by $\frac{1}{2\omega}$ in (4.6), we observe an energy drop of $10 \log_{10} \frac{1}{2} \cong -3\text{dB}$. This means that the energy spectrum of the ESS has a $-3\text{dB}/\text{octave}$ slope, as can be seen in Fig 4.6.

4.4 LINEARISATION OF THE IMPULSE RESPONSE

Let $r(t)$ be the room/cavity response to the excitation signal $s(t)$ defined in (4.1). The room/cavity impulse response $h(t)$ can be extracted by convolving $r(t)$ with the inverse filter of $s(t)$ [34] [35] [64]. The exponential sweep (which is a causal signal) is temporally reversed and then delayed to obtain a causal system [44]. However, if we time-reverse the excitation signal $s(t)$ (see Fig 4.7), it still exhibits a $-3\text{dB}/\text{octave}$. Therefore, we need to compensate this energy drop by modulating the amplitude of the time-reversed signal with a $+6\text{dB}/\text{octave}$ envelope so that the inverse filter exhibits a $+3\text{dB}/\text{octave}$ slope [34] [64]. Let us create an inverse filter $f(t)$ so that, after being convolved with the system response, it yields to the impulse response.

$$h(t) = r(t) * f(t) \quad (4.7)$$

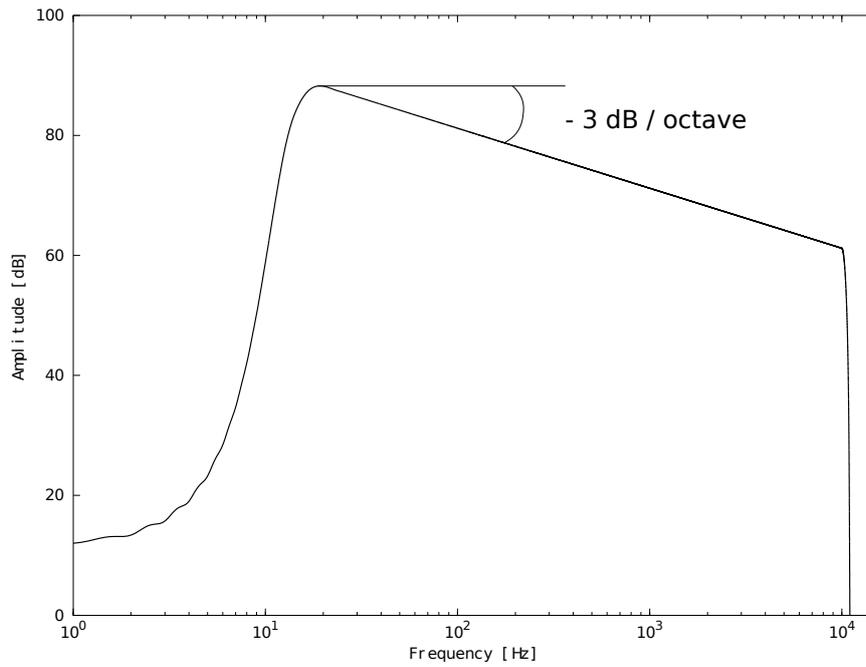


Figure 4.6: The spectrum of the exponential sine sweep displays an energy drop of -3dB/octave

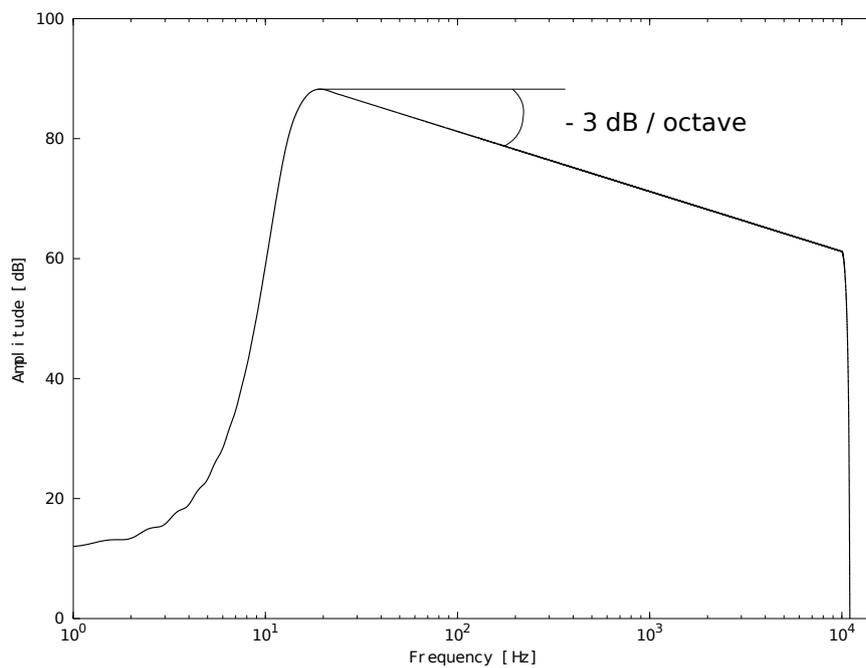


Figure 4.7: The spectrum of the time-reversed exponential sine sweep displays an energy drop of -3dB/octave

This is termed post-modulation, in opposition to a pre-modulation suggested by [64], which modulates the input signal directly so that it has a flat spectrum and the reversed-time signal also exhibits a flat spectrum. The pre-modulation necessitates adjustments prior to the recording. Therefore the robustness of the post-modulation was chosen for this experiment. The form of the post-modulation is [64]:

$$m(t) = \frac{A}{\omega(t)} = A \left[\frac{K}{L} \cdot e^{\frac{t}{L}} \right]^{-1} \quad (4.8)$$

where A is a scalar representing the modulation amplitude. At time $t = 0$, the instantaneous frequency ω equals ω_1 . In this condition, we can solve for A in (4.8), assuming arbitrarily that $m(t) = 1$ at $t = 0$:

$$m(0) = \frac{A}{\omega(0)} \rightarrow 1 = \frac{A}{\omega_1} \rightarrow A = \omega_1$$

from which we can now write (4.8) as

$$m(t) = \frac{\omega_1}{\omega(t)} = \omega_1 \cdot \left[\frac{K}{L} \cdot e^{\frac{t}{L}} \right]^{-1}$$

Modulating the time-reversed signal gives:

$$f(t) = \frac{\omega_1}{\omega(t)} \cdot \sin[\theta(T-t)] = \omega_1 \cdot \frac{L}{K} \cdot e^{-\frac{t}{L}} \cdot \sin\left[K \cdot \left(e^{\frac{T-t}{L}} - 1\right)\right] \quad (4.9)$$

and exhibits a slope of +3dB/octave as can be seen in Fig 4.7.

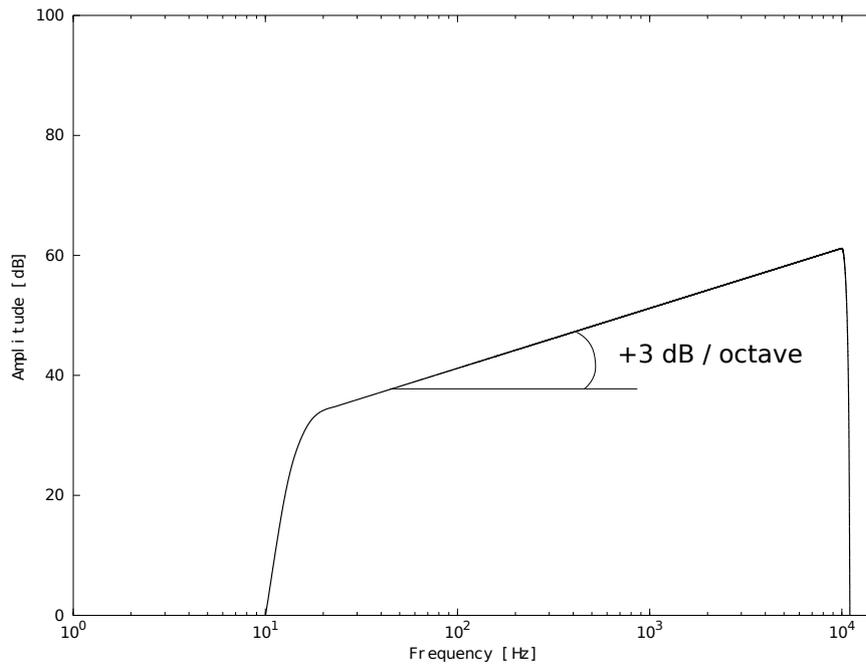


Figure 4.8: The spectrum of the amplitude-modulated exponential sine sweep displays an energy rise of +3dB/octave

Now having designed an inverse filter which counter-balances the -3dB/octave , it has to be convolved with the system response. The convolution results in a series of impulse responses, separated on the time axis. As can be seen in Fig 4.9, the Linear Impulse Response (LIR) of the system and its harmonic distortions are temporally separated on the time axis. Hence, access can be gained simultaneously to the LIR and the impulse response of each harmonic distortion.

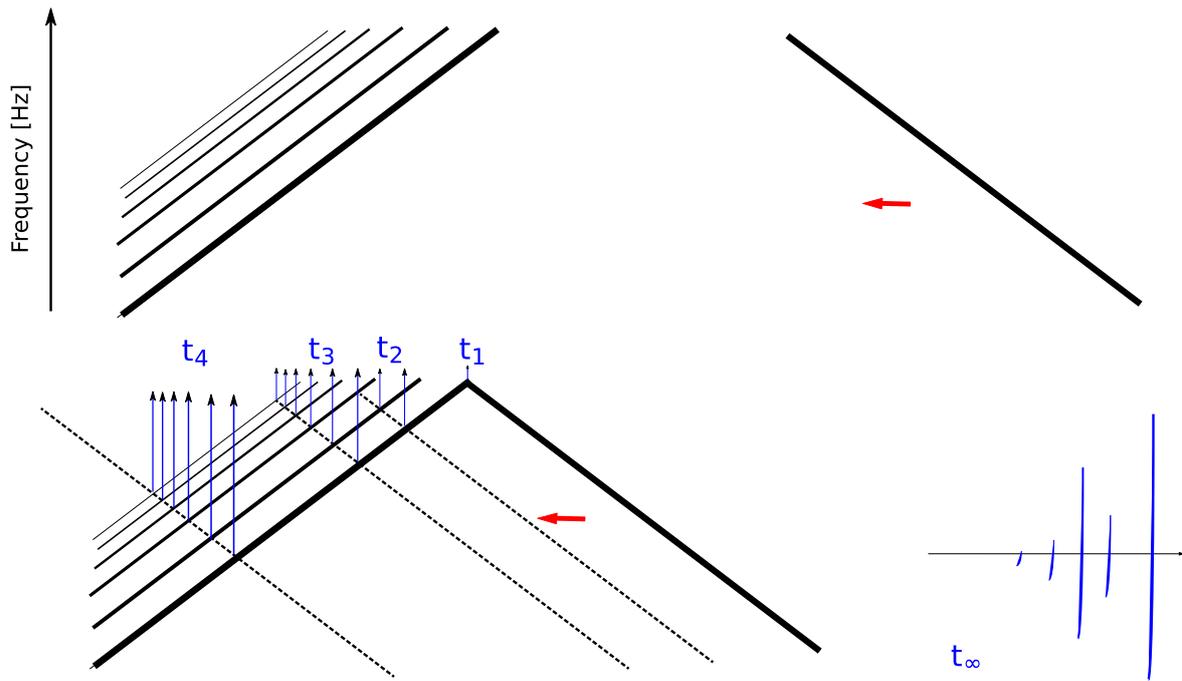


Figure 4.9: Convolution of the system response with the inverse filter signal.

The plots are spectrograms: the horizontal axis is the time and the vertical axis is the frequency. On the upper left part, the system response is plotted (the thick line is the linear response of the system and the parallel lines are the harmonic distortions). On the upper right part is the inverse filter of the input signal. Both are convolved in the time domain (horizontal axis). The convolution process is shown at different times (t_1, t_2, t_3, t_4) on the bottom part of the figure: the intersections of the system response with the inverse filter leads to the LIR split from its harmonic distortions (bottom right).

4.4.1 Note about the harmonic distortions

Electro-mechanical transducers, such as those used in speakers and microphones, are non-linear systems, i.e. they do not react proportionally to the input signal with which they are provided. On top of the linear response of the system, the transducer resonates at several frequencies; the harmonics of the linear response. These harmonic

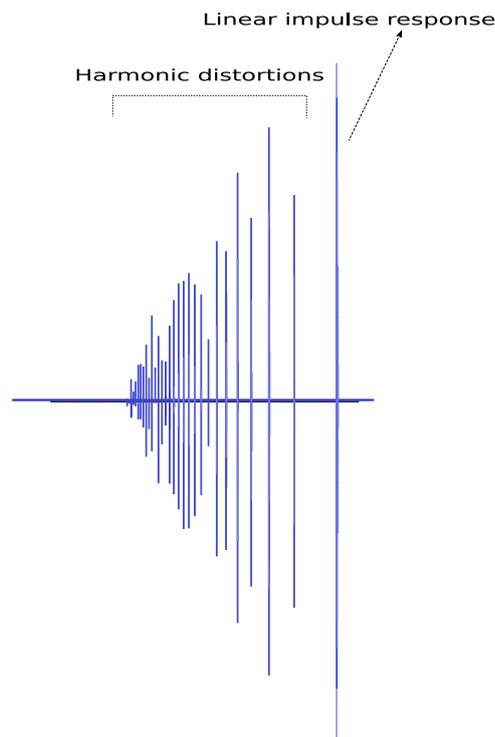


Figure 4.10: Result after the convolution of the system response with the inverse filter: the harmonic distortions are packed before the linear response on the time axis, parallel to the latter. Scale is in dB to enhance visualisation.

are called *harmonic distortions* and are inherent to every transducer. Therefore, the method described in 4.4 allows access to the linear response deprived from the harmonic distortions generated in both the speaker and the microphone. Therefore, this method is essentially independent of the speaker and the microphone.

The convolution packs the harmonic distortions before the linear response on the time axis, parallel to the latter, as can be seen on Fig 4.10. The linear response is situated at the time duration of the sweep and the harmonic distortions are parallel to it. Δt in (4.4) gives the distance on the time axis between the linear response and the N^{th} harmonic.

The big improvement with the method developed in [34] resides in the fact that applying a Fast Fourier Transform (FFT) to the Linear

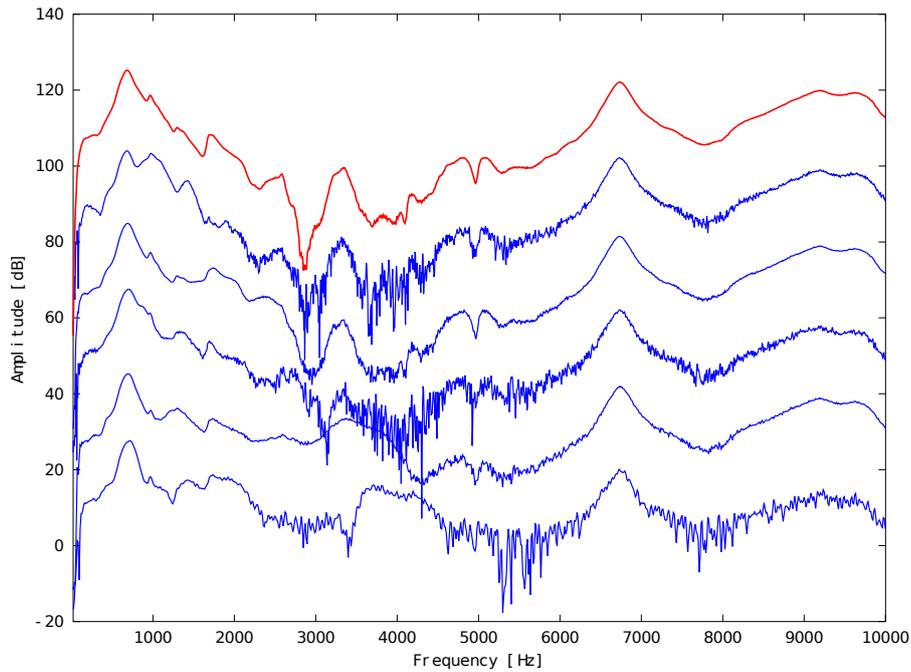


Figure 4.11: Spectrum of the LIR (in red) and its harmonic distortions (in blue).

Impulse Response or the harmonic distortions of the system allows a clear image of the transfer function of the system in the frequency domain to be seen, both for the linear response and for the non-linear distortions behaviour.

4.5 FAST FOURIER TRANSFORM

Each impulse response, starting with the LIR, is manually isolated from the other impulse responses and an FFT (Fast Fourier Transform) is performed on it, leading to the linear transfer function of the system. In Fig 4.11, one can see the transfer function of each harmonic, including the linear response.

To isolate the LIR, the software Audacity was used to zoom onto a window encompassing only the linear response, then the amplitude was switched to a logarithmic scale to assess more accurately where the impulse response starts and ends.

To perform the FFT, an algorithm (shown in Fig 4.12) was used on each impulse response p of time duration l , representing a series of recorded pressure taken at each $1/f_s$ second, where f_s represents the sampling rate:

1. **Find the next power of 2**

$$a : 2^a \geq l$$

2. **Normalisation**

$$p = \frac{p}{\max|p|}$$

3. **Zero-padding**

$$p(1 : \text{round}((2^a - 1)/2)) = 0,$$

$$p(2^a - \text{round}((2^a - 1)/2) : 2^a) = 0$$

4. **FFT**

$$20 \cdot \log_{10} |\text{FFT}(p, 2^a)|$$

This process is realised 5 times per sound sample and averaged in order to clear the inherent noise.

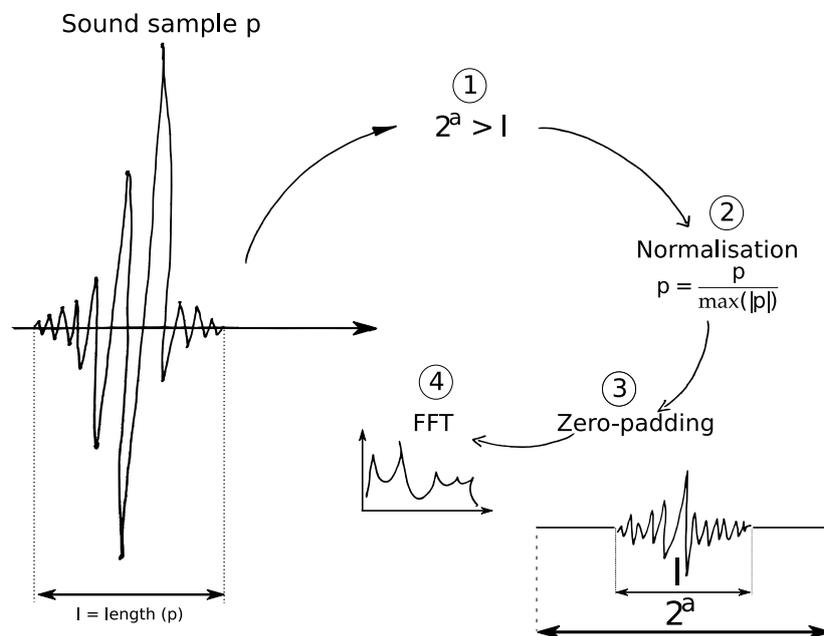


Figure 4.12: Algorithm used to obtain the transfer function out of an impulse response.

4.6 FINAL TRANSFER FUNCTION

As described in the introduction section 4.1, in Fig 4.1, we need to perform the processes 1 to 4 twice, once to obtain the transfer function of the VTM + point source (from the driver), once to obtain the transfer function of the point source from the driver alone. We can then subtract both spectra to get the transfer function of the VT model, independently from the driver.

4.7 NOTE ABOUT ESS AND ESS^{-1}

Using the ESS (4.1) as an input signal, the inverse filter (4.9), and plotting spectrograms (frequency versus time), it can be seen that there is an instantaneous burst of energy at the start and at the end of the sweep (see the green vertical lines in Fig 4.15). These are due to the fact that the sweep starts and ends non-smoothly, i.e. the slope is not continuous at the time $t = 0$ and the sweep does not necessarily cross the time axis at $t = T$. If we convolve both those signals we end up with an impulse response and its echoes in the frequency-time space, as in Fig 4.15. The idea is to provide the sine sweep with a fade-in and a fade-out.

4.7.1 A smooth start

Fig 4.13 shows that the transition at the start of the sweep is not smooth. This is due to the fact that before the sweep, the signal has a constant zero value, with a zero slope, and suddenly, at the start of the sweep, the slope is different than zero, there is no continuous transition and this results in a burst of energy on the whole spectrum, preceding the sweep.

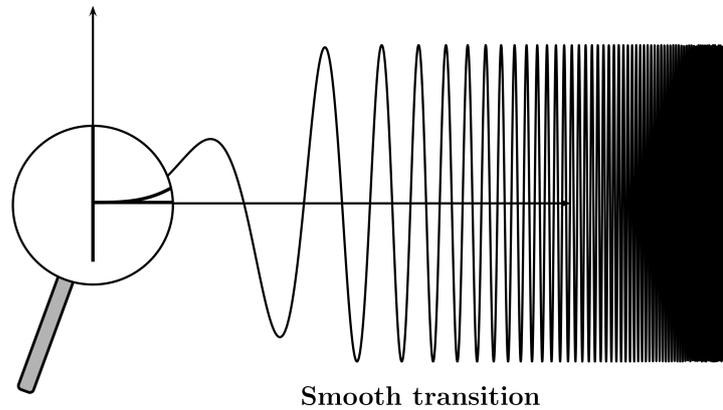
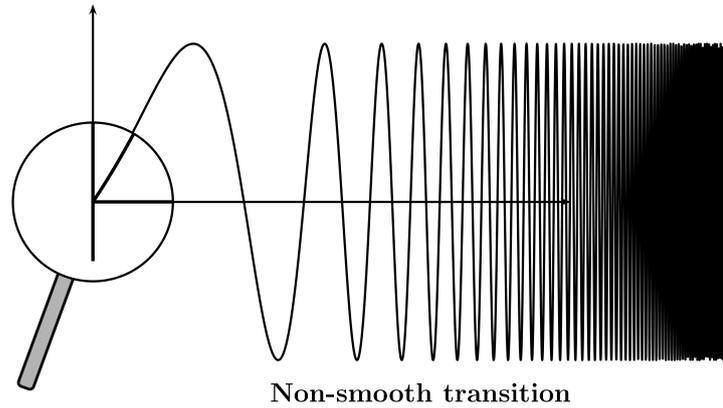


Figure 4.13: The start of the sweep is not smooth. After applying the envelope described in (4.10) and (4.11), the signal goes smoothly from zero to the start of the sweep.

The first derivative at the time origin gives the transition slope. The first time derivative of (4.1) is

$$\begin{aligned}
 \left. \frac{d[s(t)]}{dt} \right|_{t=0} &= \left. \frac{d[\sin[\theta(t)]]}{dt} \right|_{t=0} \\
 &= \frac{K}{L} \cdot e^{\frac{t}{L}} \cdot \cos \left[K \cdot \left(e^{\frac{t}{L}} - 1 \right) \right] \Big|_{t=0} \\
 &= \frac{K}{L} \\
 &= \omega_1
 \end{aligned}$$

which gives a slope different from zero.

To smooth the transition, the start of the signal was multiplied by a squared-sinus envelope (the result is displayed in Fig 4.13). Being

part of the sigmoid family, it ensures a smooth transition between a threshold value and a fixed value. This transition is applied between the start frequency of the sweep, f_1 and ends at a frequency fixed by the user, f_{in} . The algorithm works as such:

1. Find the time at which the instantaneous frequency is equal to

$$f_{in}.$$

$$t_{in} = L \cdot \ln(2\pi f_{in} \cdot L/K)$$

2. Find the maximum sampled time lesser than or equal to t_{in} .

$$t[in] = \max(t[in] \leq t_{in})$$

3. Generate the envelope.

$$\sin^2(at + b)$$

4. Multiply the signal by the envelope from $t = 0$ to $t = t[in]$.

The purpose is to find a formula of a squared sinus such that it starts at zero at the frequency f_1 and reaches the value 1 at the frequency f_{in} , after a quarter of a period. In other words, find parameters a and b such as

$$\sin^2(at + b) \Big|_{t=0} = 0 \quad \Rightarrow \quad b = 0 \quad (4.10)$$

$$\sin^2(at + b) \Big|_{t=t[in]} = 1 \quad \Rightarrow \quad a = \frac{\pi}{2t[in]} \quad (4.11)$$

Once the pre-envelope has been applied, we see that the left vertical green line (the broad-band burst of energy preceding the sweep), the "pre-ringing", to quote Farina [34] disappears as shown on Fig 4.15.

4.7.2 A smooth end

The sweep stops abruptly as soon as the frequency upper limit has been reached (see Fig 4.14). And it is very unlikely that this frequency stops when the sinus crosses the axis. For this reason, the sine sweep

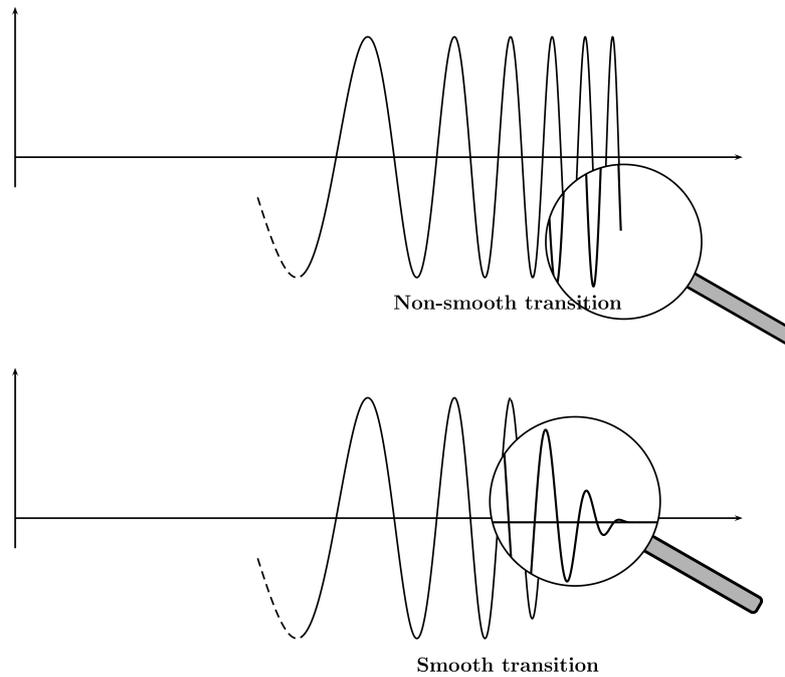


Figure 4.14: The sweep does not stop smoothly. After applying the envelope described in (4.10) and (4.11), the signal goes smoothly from the end of the sweep to zero.

defined in (4.1) generally creates a broad-band burst of energy, occurring at its end. Similarly to the method applied for the pre-envelope, a post-envelope needs to be performed to smooth down the end of the sweep onto zero. For this purpose, we apply a squared sinus which takes the value 1 at an upper fixed frequency f_{out} and goes smoothly to zero at f_2 .

The algorithm works as follows:

1. Find the time at which the instantaneous frequency is equal to

$$f_{out}.$$

$$t_{out} = L \cdot \ln(2\pi f_{in} \cdot L/K)$$

2. Find the minimum sampled time greater than or equal to t_{out} .

$$t[out] = \min(t[out] \geq t_{out})$$

3. Generate the envelope.

$$\sin^2(at + b)$$

4. Multiply the signal by the envelope from $t = t_{\text{out}}0$ to $t = T$.

We need to find parameters a and b such as the squared sinus goes from the value 1 at $t = t_{\text{out}}$ to zero at $t = T$

$$\sin^2(at + b) \Big|_{t=t_{\text{out}}} = 1 \quad \Rightarrow \quad at_{\text{out}} + b = \frac{\pi}{2} \quad (4.12)$$

$$\sin^2(at + b) \Big|_{t=T} = 0 \quad \Rightarrow \quad aT + b = 0 \quad (4.13)$$

Subtracting (4.13) from (4.12) gives

$$a = \frac{\pi}{2(T - t_{\text{out}})} \quad (4.14)$$

Once the pre- and post-envelopes have been applied, we see that both the left and the right vertical green line (the broad-band burst of energy preceding and following the sweep respectively), the "pre-ringing" and the "post-ringing" [34] disappear as shown on Fig 4.15.

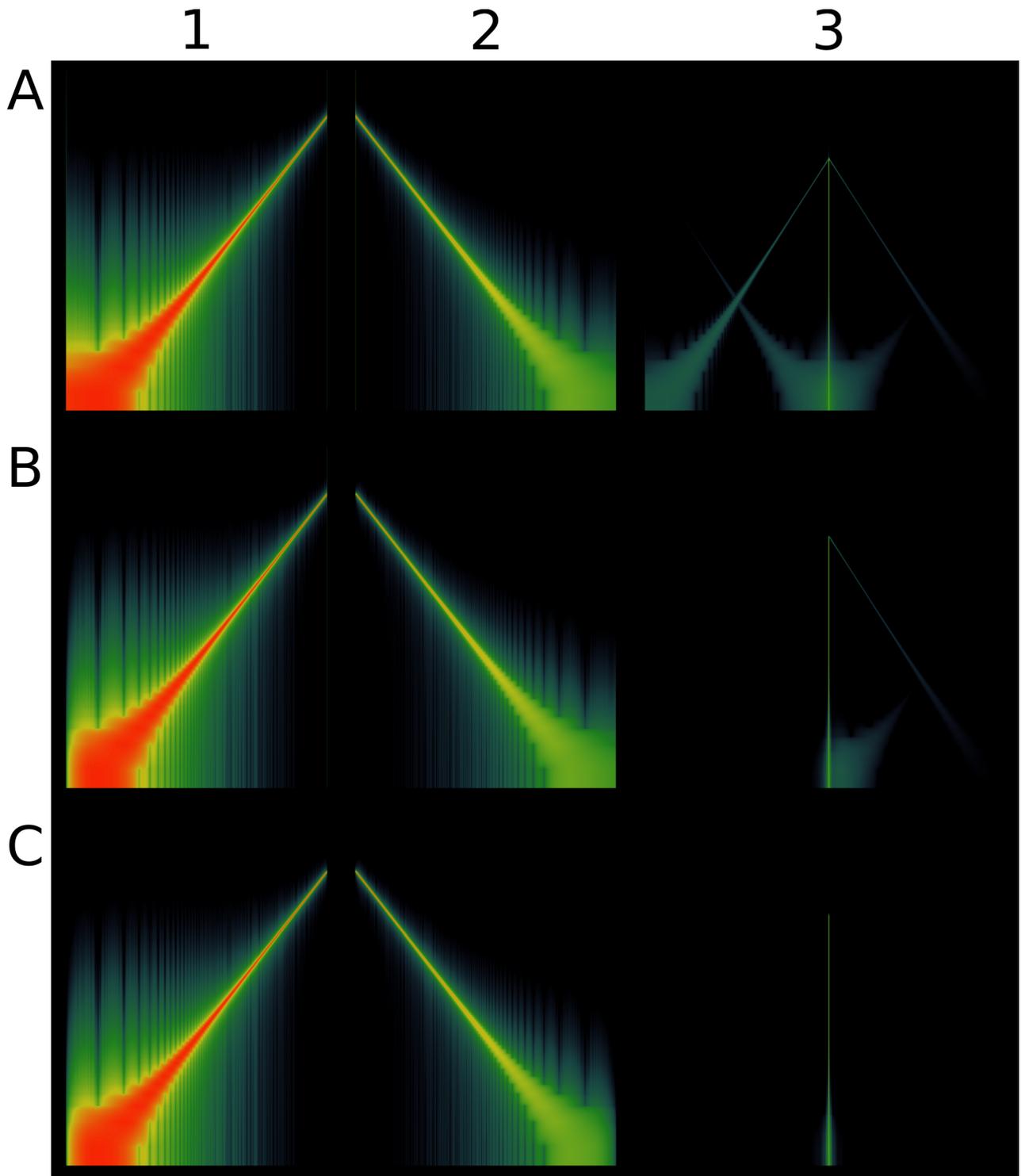


Figure 4.15: An Exponential Sine Sweep (ESS) of the form (4.1) has a burst of energy across the whole spectrum both at its start and at its end (A₁). Once convolved with its inverse filter (A₂), it leads to an impulse response and its echoes in the frequency-time space (A₃). Providing a smooth start to the (ESS) (B₁), and convolving it with its inverse filter (B₂) removes the pre-ringing (B₃). Providing the (ESS) with both a smooth start and a smooth end (C₁), and convolving it with its inverse filter (C₂) removes both the pre- and the post-ringing (C₃).

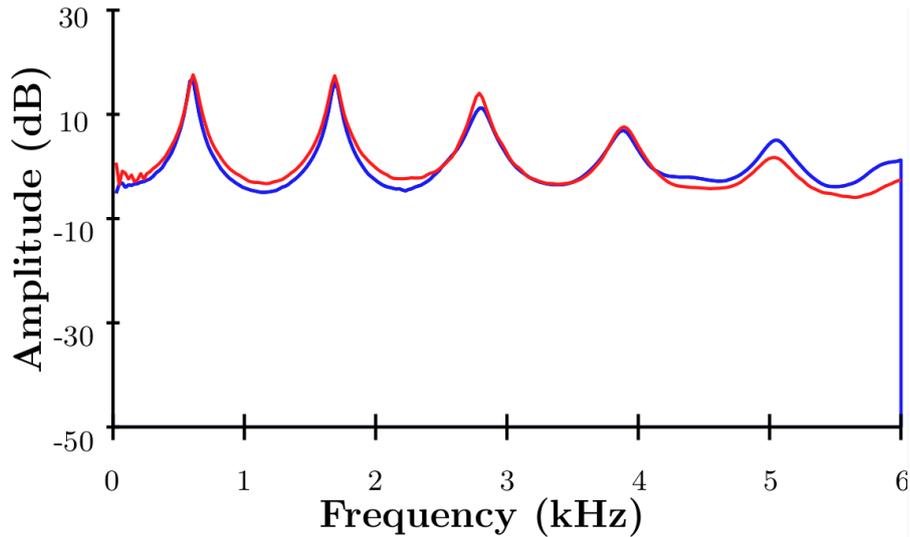


Figure 4.16: Driver-independent experimental method. Measured transfer functions of the resonance of VTM-1 with 2 different speakers: a 952.210UK driver unit (in blue) and a VISATON SC 8 N 8 Ω transducer (in red).

4.8 BENCHMARKING

This section illustrates the driver-independence of the experimental method and the implications of changing the distance of the probe microphone in the numerical simulations as well as the absorption coefficient.

4.8.1 Driver-independent

On Fig 4.16, one can see that the experimental method is essentially driver-independent. The resonances of a single tube (VTM-1) of Length = 142 mm, Radius = 15 mm, Flange = 2 mm were excited by 2 inexpensive drivers at the lips end: a 952.210UK driver unit (in blue) and a VISATON SC 8 N 8 Ω transducer (in red). Both drivers show very similar results, making the experimental method essentially speaker-independent.

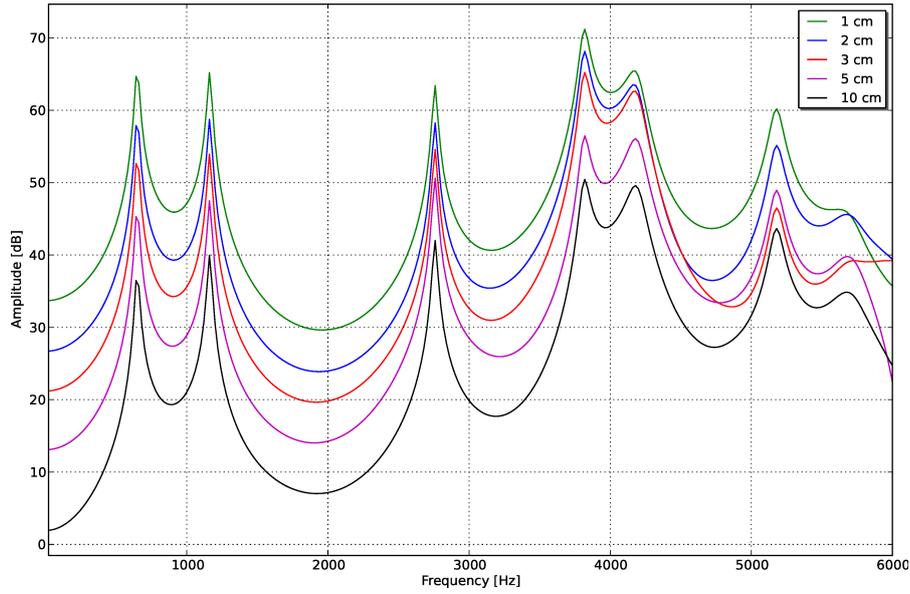


Figure 4.17: Distance-dependence of the resonances of VTM-Ch&K-/a/. Simulated transfer functions of the resonances of VTM-Ch&K-/a/, when the distance between the lips opening and the virtual probe takes the values 1cm, 2cm, 3cm, 5cm and 10 cm.

4.8.2 Distance dependence

Figs 4.17 and 4.18 illustrate the distance dependence of the numerical method for the probe location for VTM-Ch&K-/a/ and VTM-MRI-Barnaby-/hard/ respectively. The further the probe is from the lips the more the overall amplitude of the spectrum decreases, as expected from a sound source radiation, the radiated power of which decreases as the inverse of the squared distance. The radiation propagates better the high frequencies, as seen in subsection 2.3.1.

4.8.3 Absorption coefficient dependence

To simulate the walls' reflections and the absorbent behaviour of the actual material (and of the VT walls) in the numerical method, a normal acoustic admittance is applied on the walls of the VT models. This is performed by a simple method given in [121], see chapter 3. The acoustical absorption coefficient is given as

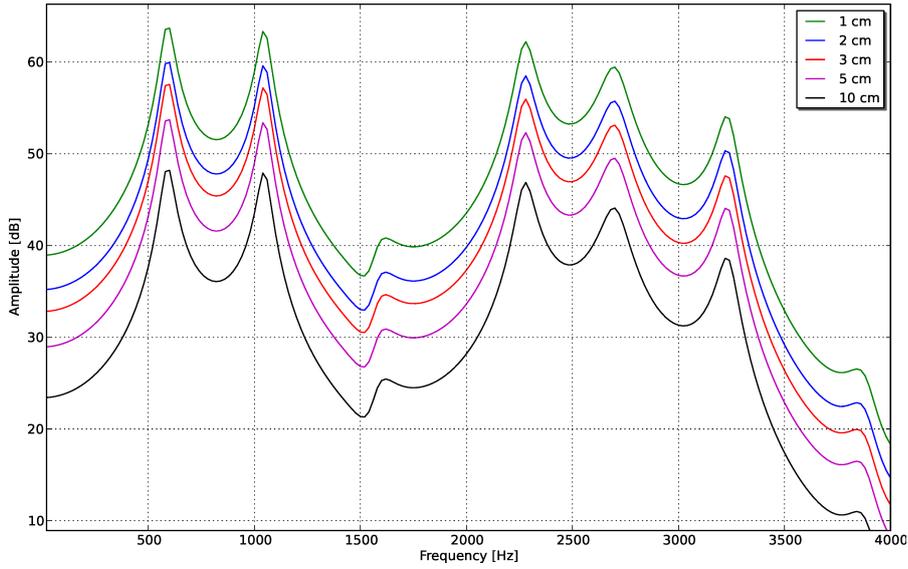


Figure 4.18: Distance-dependence of the resonances of VTM-MRI-Barnaby-/hard/. Simulated transfer functions of the resonances of VTM-MRI-Barnaby-/hard/, when the distance between the lips opening and the virtual probe takes the values 1cm, 2cm, 3cm, 5cm and 10 cm.

$$\alpha = 1 - \left| \frac{Z_n / (\rho c) - 1}{Z_n / (\rho c) + 1} \right|^2 \quad (4.15)$$

where α is the absorption coefficient, Z_n the normal acoustic impedance, ρ the density of air, c the speed of sound. (4.15) leads to

$$Z_n = \rho c \frac{1 + \sqrt{1 - \alpha}}{1 - \sqrt{1 - \alpha}} \quad (4.16)$$

From 4.16, the normal admittance can be derived as

$$A_n = \frac{1}{Z_n} \quad (4.17)$$

In the conditions of the experiment ($T = 5^\circ$), the different values can be computed:

Absorption	
α	A_n
0.001	6.1075×10^{-7}
0.01	6.03226×10^{-6}
0.03	1.8593×10^{-5}
0.1	6.4301×10^{-5}

The results can be seen on Figs 4.19 and 4.20, which display the transfer functions with the different absorption coefficients from table 4.8.3 for VTM-Ch&K-/a/ and VTM-MRI-Barnaby-/hard/ respectively. It can be seen that the more absorbent the VT walls are, the more the amplitude of all the formants decrease. This leads progressively to the assimilation of F₄-F₅ and quasi assimilation of F₁-F₂ in Fig 4.19, and the assimilation of F₁-F₂ and F₃-F₄-F₅ in 4.20 for the most absorbent coefficient shown ($\alpha = 0.1$). It is observed that a greater absorption coefficient of the Vocal Tract walls tends to cluster the formants together.

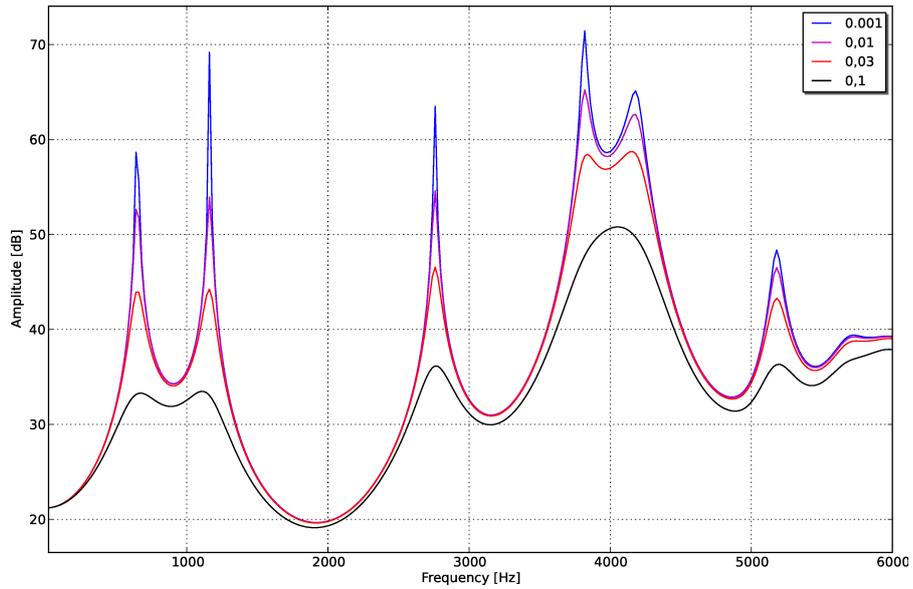


Figure 4.19: Absorption-dependence of the resonances of VTM-Ch&K-/a/. The simulated transfer function of VTM-Ch&K-/a/ is plotted for different values of absorption coefficient ($\alpha = 0.001, 0.01, 0.03, 0.1$).

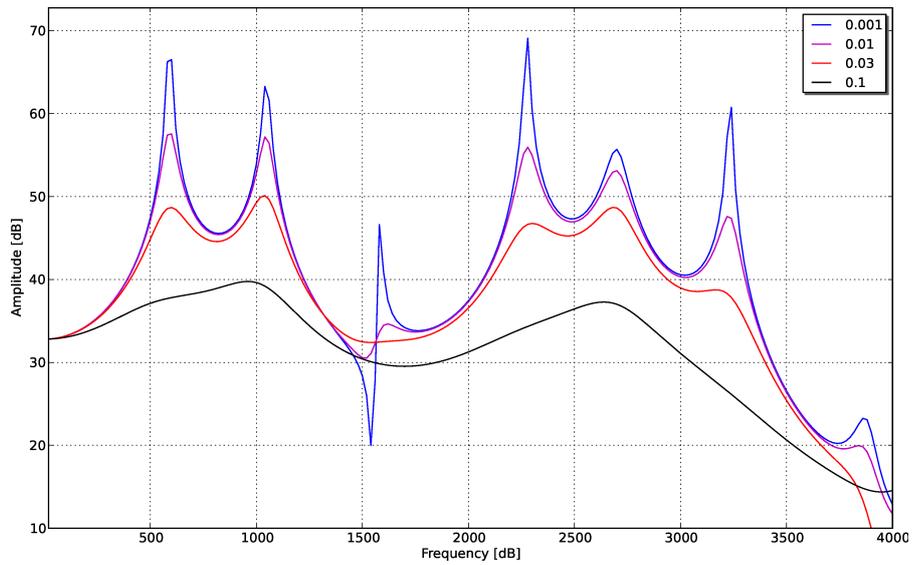


Figure 4.20: Absorption-dependence of the resonances of VTM-MRI-Barnaby-/hard/. The simulated transfer function of VTM-MRI-Barnaby-/hard/ is plotted for different values of absorption coefficient ($\alpha = 0.001, 0.01, 0.03, 0.1$).

5 | FROM MRI TO 3D PRINTING ...

The previous chapter described the material and method used to measure the transfer functions of the different models of the Vocal Tract. This chapter explains the methodology employed from the data collection of a singer phonating in an MRI scanner to the measurement of their 3D-printed tract.

5.1 MAGNETIC RESONANCE IMAGING

Magnetic resonance imaging has been used to acquire 3D representation of the head and neck of a singer phonating. This allows to reconstruct the shape of the VT during phonation. This shape will be used either for rapid prototyping (3D printing) or as a basic computational domain for numerical simulation using the FEM. For an extensive development on MRI technique, refer to Appendix D. Obtaining optimal results with the MRI technique requires a compromise between minimal capture time (to reduce motion artefacting) and maximal resolution: a single image can be captured in the midsagittal plane in under a second, but a 3D capture necessitates parallel slices acquisition and requires the subject to remain stationary for the duration of the scan [86].

5.1.1 Acquisition and protocol

As mentioned in section ii, data from six professional singers were acquired in two research facilities, the "York" and "Freiburg" protocols respectively.

5.1.1.1 Protocol "York"

The scans of Maristela, Bartholomew and Barnaby (pseudonyms for the singers, see section ii) have been acquired according to the protocol described in [86]. In this study, participants were asked to phonate on a phoneme (see Table 2.2) for as long as comfortably possible, and instructed to then attempt maintaining the articulatory setting in an unvoiced condition whilst breathing for the remainder of the scan [86]. Scans are made at the York Neuroimaging Centre (YNiC), using a General Electric 3.0 T HDx Excite MRI Scanner. The scan developed was a 3D fast gradient echo sequence, the details of which are summarised in Table 5.1. From [85]: *Acquisition is isotropic 2mm in a 192×192 matrix. Output is then interpolated to 512×512 using 50% slice overlap giving an effective anisotropic output of $0.75 \times 0.75 \times 1$ mm. A stack of 80 images is produced in the midsagittal plane in approximately 16 s.* Maristela and Barnaby's MRI data were acquired by Matt Speed [85] whilst Bartholomew's MRI data was acquired by the author of this thesis.

5.1.1.2 Protocol "Freiburg"

The other set of MRI data i.e. Sophy, Marisa and Timothy were acquired in the Department of Radiology of the University Medical Center Freiburg, in Germany. The subjects were examined radiologically with the 3.0 T TIM TRIO (Siemens, Germany) MRI. The subjects sustained a tone for 20 s, sung on the vowel /a/ (as in /hard/, see Table 2.2) in modal register; they were asked to sustain the tone as consistently as possible [27]. The data for the acquisition: an effective

“York” Protocol

T_e	1.7ms
T_R	4.8ms
Flip Angle	5°
Bandwidth	$\pm 41.67\text{Hz}$
FOV	260mm ³
Slice Width	2mm (50% separation)
Matrix	512 × 512

Table 5.1: “York” protocol: technical data for the MRI data acquisition of Maristela, Bartholomew and Barnaby.

anisotropic output of $1 \times 1 \times 1.3\text{mm}$ in the midsagittal plane acquired in approximately 12.93s (see Table 5.2).

5.1.1.3 *Quality of MRI data*

Limitations in MR image resolution and accuracy [5] (in section i) as well as air-to-tissue boundaries distortions due to MRI artefacts [91] can lead to the blurring of the edges. The protocol “Freiburg” has a resolution of 224×256 versus a 512×512 matrix for the protocol “York”. Therefore, the image acquired with the latter are of a better resolution and accuracy than the former. Artefacts due to movements during phonation for Maristela [85] and Marisa also blurred the pictures, and made them more difficult to segment accurately. So the quality of the acquired MRI data can be ranked by order of increasing quality: Marisa, Sophy, Timothy, Maristela, Bartholomew, Barnaby.

5.1.2 Segmentation

The MRI acquisition gives a stack of 2D images, from which the volume of the tract is reconstructed in 3D. For that purpose, ITK-Snap

"Freiburg" Protocol

T_e	1.67ms
T_R	4.85ms
Flip Angle	12°
Bandwidth	$\pm 123\text{Hz}$
FOV	$260 \times 227.5 \times 62.4\text{mm}^3$
Slice Width	1.3ms
Matrix	224×256

Table 5.2: "Freiburg" protocol: technical data for the MRI data acquisition of Marisa, Timothy and Sophy.

(www.itksnap.org) is used; this open-source software allows segmentation of structures in 3D medical images. It can manipulate images from MRI scan (such as .DICOM files) and perform a semi-automatic or manual segmentation. Segmenting an anatomical structure in ITK-Snap involves assigning a label to each voxel (*volumetric pixel*) in the structure.

Fig 5.1 shows the mid-sagittal slice of a singer phonating. The process for the segmentation performed with ITK-Snap can be summed up in the following steps:

1. **Resampling** the region of interest (see Fig 5.2a).
2. **Intensity Regions:** parametrise the contrast to clearly display the borders of the region of interest (see Fig 5.2b).
3. **Bubbles:** propagate bubbles so that they swell until they meet a border (the different steps are illustrated on Figs 5.3).
4. **Manual Editing:** necessary post-processing after the semi-automatic segmentation, to clean the segmentation of bubbles leaking, teeth, etc (see Fig 5.4, pre- and post- manual editing).

The segmented structure can now be exported as an .STL file (*Stereolithography*, see subsection 5.2.1) for further processing (see Fig 3.3), e.g.

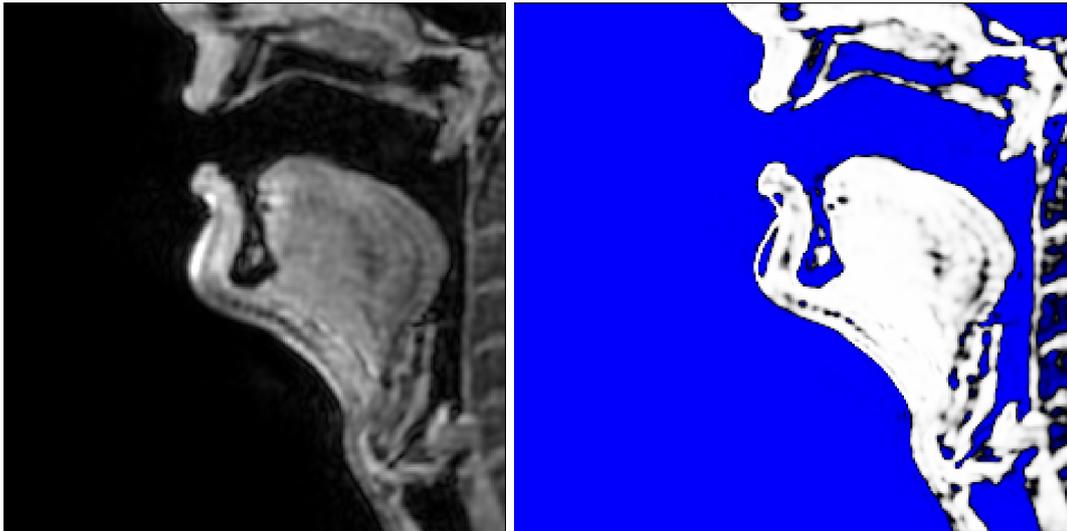
- Meshing for FEM simulation
- 3D printing

5.1.2.1 *Note about the teeth*

The proton density of the teeth makes them appear as air on MRI pictures, so that special care needs to be taken to remove the teeth from the vocal tract, by manually editing the semi-automatic segmentation. On Fig 5.5, it can be seen that the teeth appear as air in the Vocal Tract and need to be removed.



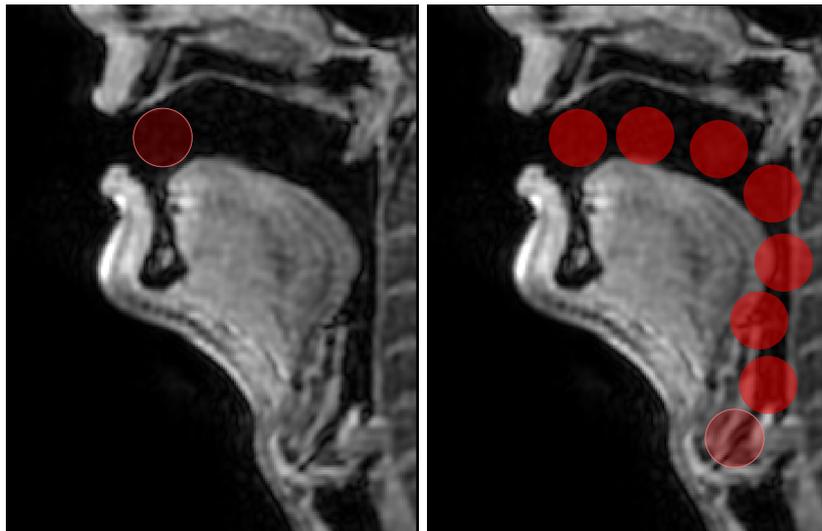
Figure 5.1: MRI mid-sagittal slice.



(a) Sagittal slice: pre-segmentation

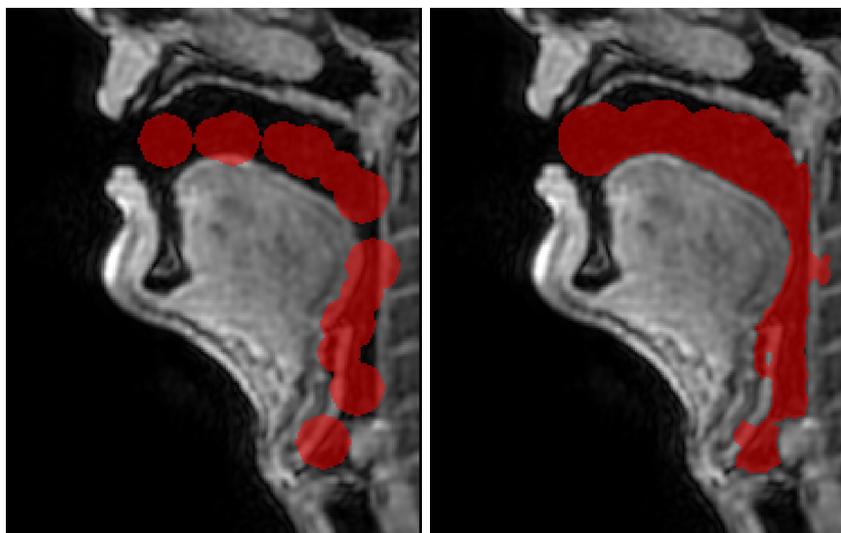
(b) Sagittal slice: intensity regions

Figure 5.2: MRI sagittal slice pre-segmentation and intensity regions.



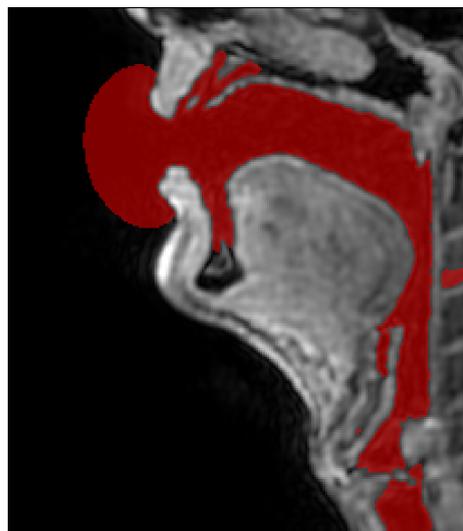
(a) First bubble

(b) Bubbles



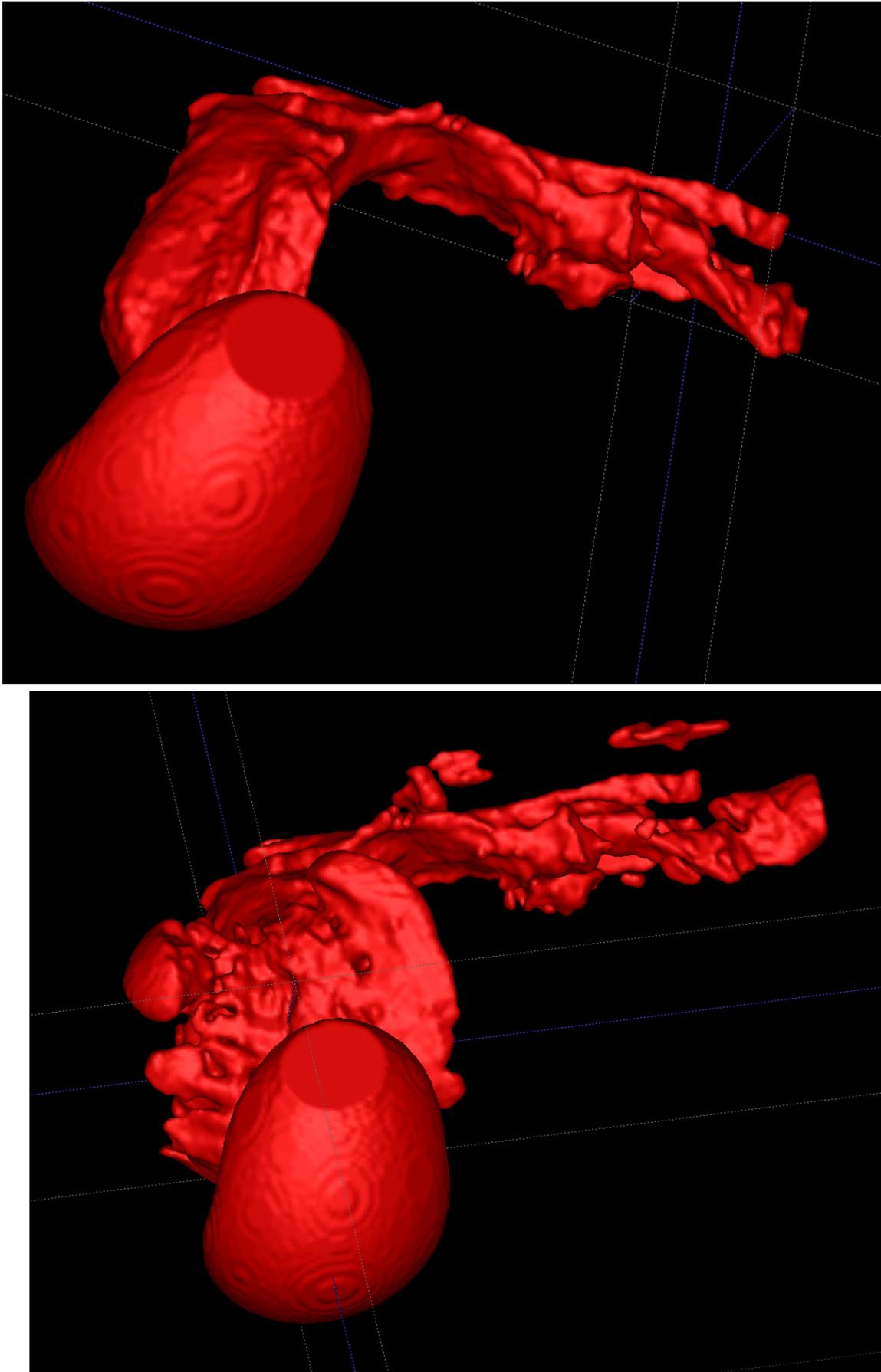
(c) Bubble expansion: start

(d) Bubble expansion



(e) Bubble expansion: end

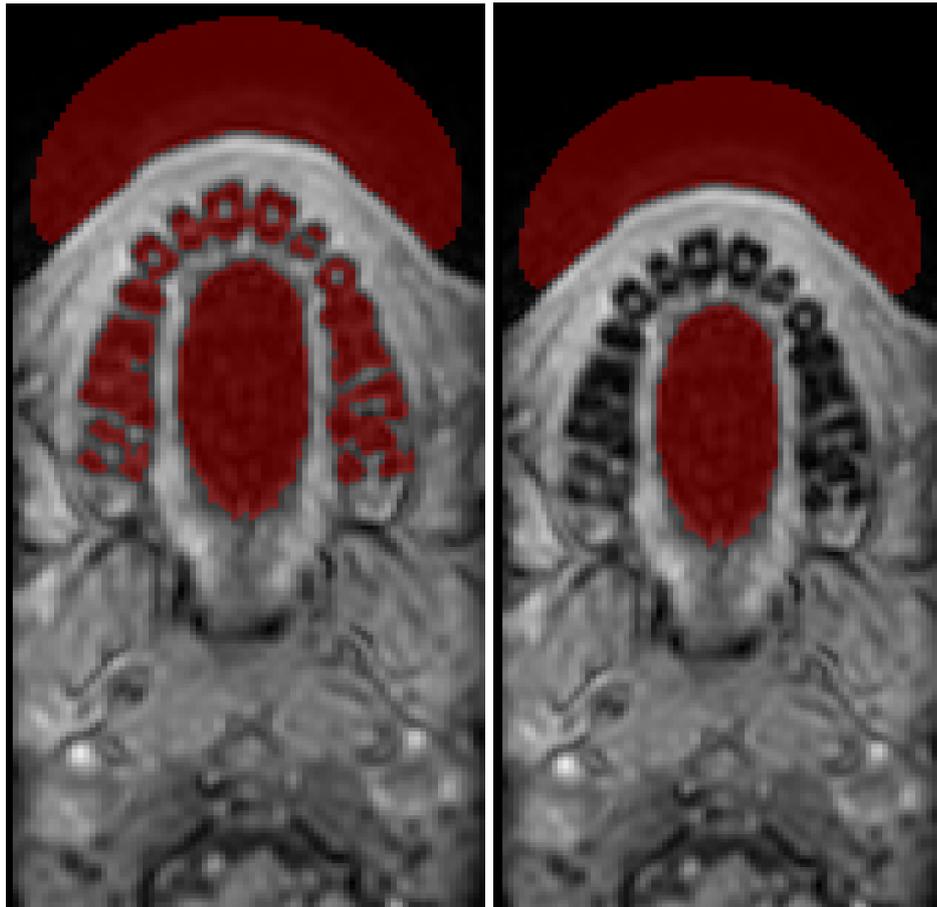
Figure 5.3: Bubbles expansion process: bubbles are placed in the vocal tract and expand until they reach a boundary defined by the intensity region.



(a) Volume representation of the segmentation, pre-editing

(b) Volume representation of the segmentation, post-editing

Figure 5.4: Manual editing after semi-automatic segmentation.



(a) Segmentation editing: with teeth

(b) Segmentation editing: without teeth

Figure 5.5: Manual editing for the teeth

5.2 3D PRINTING

3D printing provides the opportunity to mould MRI-based VTs and subsequently measure the transfer function of the tract *in vitro*. Previously, measurement was performed *in vivo* [45, 41], which is indubitably invasive and necessitates the glottis of the subject to be closed, in order not to account for any subglottal resonances and/or acoustical coupling.

The industry of 3D printing has been expanding very rapidly since its creation in 1984 when Charles W. Hull first introduced the term *stereolithography*, a printing process that enables a 3D object to be created from a digital file. This technology is used to create a 3D model from a picture and gives the users the possibility to test a design before launching a larger manufacturing program. At the time, it was too expensive for industry for manufacturing, but 30 years later the costs of 3D printing have dropped dramatically and it has become affordable even for the private customer. Nowadays, the applications of 3D printing cover a broad range of fields, including architecture, engineering, industrial design, automotive design, aerospace, dental and medical industries, biotech, etc.

3D printing is also called *additive manufacturing* because it creates objects by adding layers of material, as opposed to *subtracting manufacturing*, which was the more traditional method used in the past to manufacture pieces for industry. *Subtracting process* is based on taking bits away from a solid (metal, wood, plastic, ...) by drilling, grinding and milling, especially with metallic components, whereas the *additive process* consists in depositing material on a platform, layer by layer, until full completion of the object to print.

5.2.1 Technical data

The objects to be printed are provided as .STL files (*Standard Triangulation Language*), which approximate a surface with triangular faces. The machine used to print the VTMs is called **Objet24 3D Printer**. Objet printers create models by jetting thin layers of printing materials on a platform, called the build tray, layering up the cross-sections of the object until the complete model is formed. This process uses two types of material:

- **Model material** which is what the finished model is made of.

- **Support material** which fills gaps and spaces in the model, supporting the overhanging features during printing, and is removed after printing

Materials used for printing models with Objet printers are made of resins, which are composed of reactive monomers and oligomers, called photopolymers, i.e. they can bind to one another when exposed to light, for instance. VeroWhitePlus Opaque is used as model material and FullCure 705 as support. The printer has a 28-micron resolution printing, which means that the particles (3D dots) used in the photopolymer stereolithography are around 28 μm in diameter. The maximum dimensions of a single model printed are:

- X axis: 233.00 mm (9.17 inches)
- Y axis: 191.00 mm (7.52 inches)
- Z axis: 148.30 mm (5.83 inches)

5.2.2 Pieces printed

The cabinet which encloses the speaker was first printed, followed by VTM-1 and VTM-2 respectively. Finally, the MRI-based VTs (VTM-MRI) were printed. Note that the VTM-Ch&K did not need to be 3D printed, as the VT replicas made by Arai [2], from Chiba & Kajiyama measurements [14] were used.

The VT, as its simplest model, can be approximated by one cylinder (VTM-1), which represents the oropharynx. Appending another tube fulfilling the role of the epiglarynx (VTM-2) allows to gain access into the spectral impact of this cavity.

The first piece under investigation (VTM-1) is a cylinder whose dimensions (height 142 mm, radius 15 mm, flange 2 mm) match the outline dimensions of VTM-Ch&K without the epiglaryngeal tube. It

	height [mm]	radius [mm]	
1 tube	142	15	
2 tubes	16	6	
	1	6	(x10)
	20	5	
	20	4	
	20	3	
	20	2	

Table 5.3: Dimensions of the different 3Dprinted pieces used to be combined to model the Vocal Tract (VTM-1 and VTM-2).

is represented in *orange* in Fig 5.6. A second tube (in *purple*) is then appended to VTM-1, whose height and inner radius can be modulated between 16 and 26 mm and 2 and 6 mm respectively. This 2nd tube represents the epilaryngeal tube which can be lengthened and/or narrowed in combination with the pieces in *purple* on Fig 5.6. The pieces in Fig 5.6 can be combined in several ways to assess the change in the transfer function when one goes from one tube (VTM-1) to two tubes (VTM-2), varying the length and the radius of the second. The dimensions of the different tubes are given in Table 5.3

The final pieces printed were the MRI-based VTs (VTM-MRI), two of which can be seen on Fig. 5.7. It exhibits a distinct insert at the base, which plugs onto the matching horn on top of the driver to play a glottal signal through the vocal tract for demonstration purposes. It is also provided with a hinge designed to allow the opening of the VT, to fill the piriform fossae or valleculae with plasticine to assess their spectral impact, for instance.

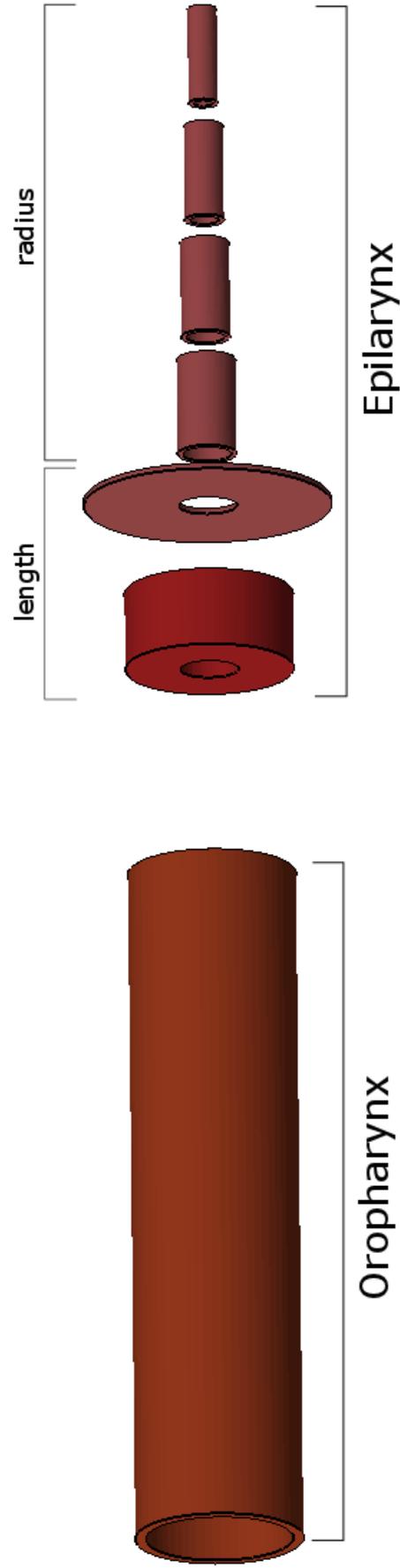


Figure 5.6: 3D printed pieces for VTM-1 (oropharynx only) and VTM-2 (oropharynx + epilarynx).

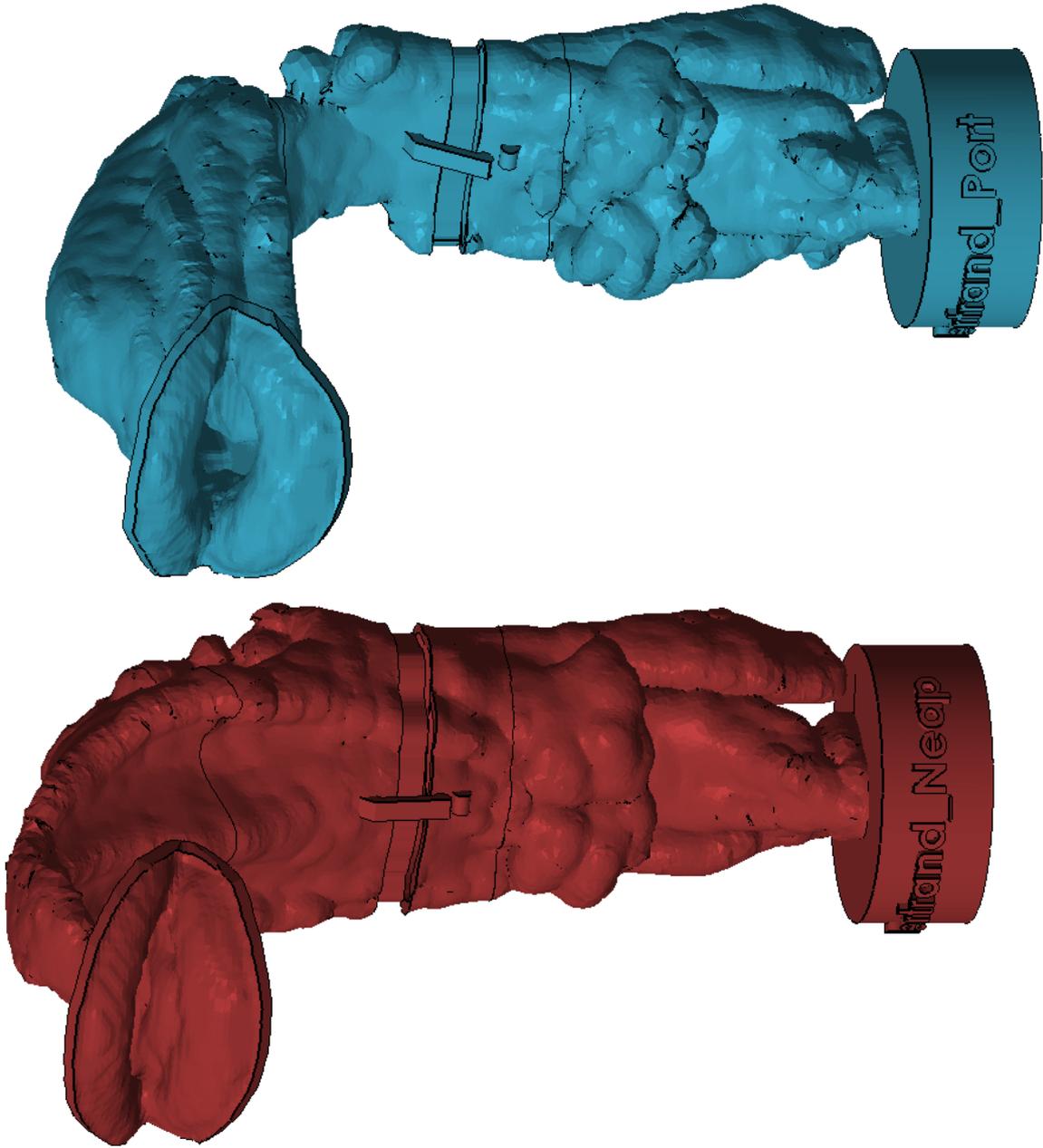


Figure 5.7: 3D printed VTM-MRI for the vowels as in /neap/ and /port/.

Part III

RESULTS

This section exposes the results obtained with the methods described in part ii. The results chapter derives the theoretical predictions of the resonances of one tube (VTM-1) and a twin-tube (VTM-2) and compares them with the numerical simulations and experimental measurements. The next section explores the relation between the dimensions (length and radius) of the epilarynx tube and the extra formant which is generated when it is appended to the oropharynx tube (from VTM-1 to VTM-2). It then switches to a more complex/realistic model of the VT, the VTM-Ch&K. Finally, the results of the most realistic VTM, i. e. VTM-MRI, are displayed with differences across vowels for the same singer (Barnaby), for SFC across registers (Bartholomew) and across singers from different voice classifications ((Sophy, Marisa, Maristela, Timothy, Bartholomew, Barnaby)). A new metric is then introduced, in order to be able to compare the SFC across singers, vowels and registers. Finally, the data from VTM-MRI are being visualised at the light of the new metrics, and a formula linking the dimensions of the hypopharynx to the SFC is suggested.

6 | RESULTS

This chapter first focuses on the spectral impact of the epilarynx, then on the spectral impact of the piriform fossae and the vallecula. It starts with the derivation of the theoretical predictions of the resonances of simple structures approximating the Vocal Tract, such as a single tube (VTM-1) or a twin tube (VTM-2). The notion of Open End Correction Coefficient is then introduced and the theoretical predictions are confronted with the experimental results and the numerical simulations. The next sections display the results of different models of the VTM, by order of increasing complexity/realism, VTM-1, VTM-2, VTM-Ch&K and VTM-MRI. A new metric is then introduced to measure the SFC. The last section visualises the data at the light of new metric.

In this chapter the results will be presented under different forms. To avoid confusion, here is a concise description of what each term refers to:

- **Theoretical predictions:** resonance frequencies of simple structures such as VTM-1 or VTM-2 analytically predicted by theory.
- **Numerical simulations:** transfer functions obtained through numerical methods (FVM or FEM) introduced in the section 3.
- **Experimental measurements:** measurements obtained in the anechoic chamber with the novel experimental method developed in section 4.

6.1 THEORETICAL PREDICTIONS

To ensure the reliability of the numerical simulations and experimental measurements, these two approaches are tested against theoretical predictions. In this section, the theoretical predictions of the resonant frequencies of one tube (VTM-1) and a twin-tube (VTM-2) are derived. The notions of acoustical length and Open End Correction Coefficient are introduced. These theoretical predictions are compared with numerical simulations. In Fant [32], the VT is approximated by a twin-tube resonator to model the different vowels whereas here the choice has been deliberately made such that both tubes act as the epiglaryngeal and the oro-pharyngeal tube, to account for the creation of the Singer's Formant Cluster (SFC).

6.1.1 VTM-1

Let us first examine a tube with a uniform cross-section. The modes of an open-closed cylinder are of the form :

$$\Psi_{n_z, m}(n) = J_m(k_{r, m}(n)r)e^{im\phi}\sin(k_z z)$$

with the eigenvalues

$$k_z = \frac{(2n + 1) \pi}{L} \frac{\pi}{2} \quad (6.1)$$

giving the name of a *quarter wavelength resonator*. (see the Appendix A for an extensive development of this solution).

When its length is large in comparison with the wavelength, the resonant frequencies of a cylinder can be approximated under the 1D assumption of plane wave propagation. The cross-sectional dimension of the tubes should be less than a half-wavelength [69], which means it is valid up until about 5 kHz (VTM-1 has a diameter of 30

mm, see section ii). Under this assumption, the acoustic modes are given as the solutions of [69, 101]

$$A \cot \frac{2\pi f L^*}{c} = 0 \quad (6.2)$$

where L^* is the acoustical length of the tube, A is its cross-section, c is the speed of sound and f is the frequency.

The resonances of VTM-1 (whose dimensions are: Length = 142 mm, Radius = 15 mm, Flange = 2mm) are displayed on Fig 6.1. The theoretical predictions are the roots of the equation (6.2), plotted in red. These are linked to the numerical simulation (FEM) in black and experimental result in grey by the dotted lines. The experimental and numerical results match the theoretical predictions of the resonant peaks in frequency.

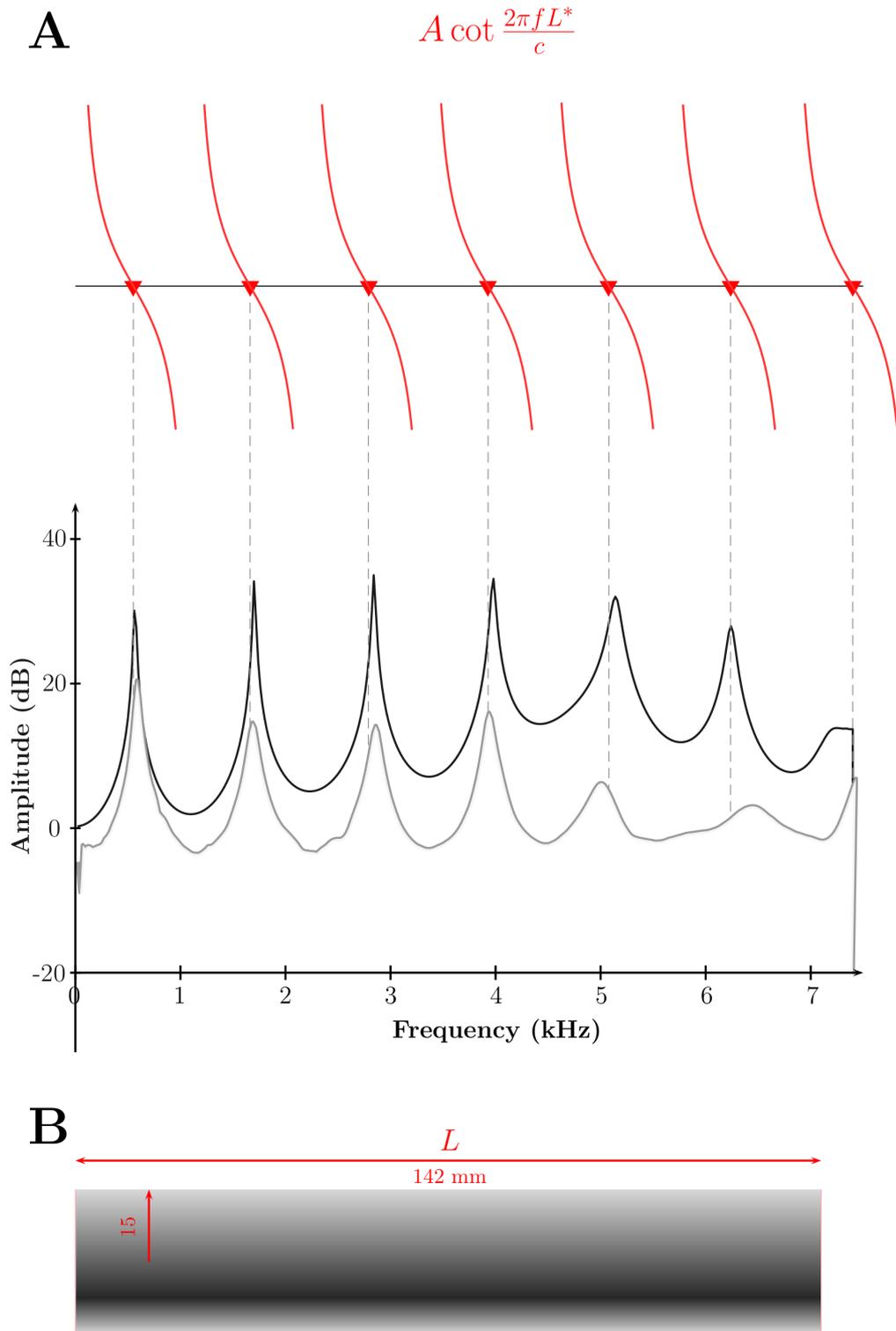


Figure 6.1: Theoretical predictions, numerical simulations (FEM) and experimental measurements of the resonances of one tube. On the upper part of **A**, the theoretical predictions are given as the roots of the equation (6.2), plotted in red. On the lower part of **A**, these are linked to the numerical simulation (FEM) in black and experimental result in grey in by grey dotted lines. **B** shows VTM-1 and its respective dimensions.

6.1.2 VTM-2

For a twin-tube resonator (VTM-2), the acoustic modes are given as the solutions [101, 57] of

$$A_1 \tan \frac{2\pi f L_1^*}{c} = A_2 \cot \frac{2\pi f L_2^*}{c} \quad (6.3)$$

which are solved graphically on Figs 6.2, 6.3 and 6.4 (A_1 and A_2 as the cross-sections of the epilaryngeal and oro-pharyngeal tubes, respectively, c is the speed of sound and f is the frequency). L_1^* and L_2^* used in the equations represent the acoustical length of the oro-pharyngeal and epilaryngeal tubes respectively.

In all the coming figures, the oro-pharyngeal tube is in red, whereas the epilaryngeal tube and its related effects are in blue. The greyscale pattern represents the model of the whole VT (blue + red). Figs 6.2, 6.3 and 6.4 illustrate the creation of an extra resonance when a second tube is appended to the cylinder of Fig 6.1 (i. e., from VTM-1 to VTM-2), via numerical simulations (FEM and FVM) and experimental results respectively. This extra resonance is synonymous with an extra formant and the creation of the SFC once the VT model is evolved into a more sophisticated model, the VTM-MRI.

Figs 6.2B, 6.3B and 6.4B show the simple tube (red) and the twin-tube resonator (blue + red). On Figs 6.2A, 6.3A and 6.4A, we can see the resonances of the simple tube, whose theoretical predictions are given by the roots of the red function (6.2) or right-hand term of the (6.3), symbolised by red triangles whereas the numerical results are plotted in red, with white labels for the resonances. In blue, the asymptote of the blue function (left-hand term of (6.3)) gives the theoretical prediction of the first resonant frequency for the standalone epilaryngeal tube. The intersections between this blue function and the red function give the predicted resonance frequencies for the twin-tube resonator (black diamonds on On Figs 6.2A, 6.3A and 6.4A). The corresponding numerical (experimental) results are plotted in

greyscale. The dashed black lines link the theoretical predictions and the numerical (experimental) results for the formant frequencies of the twin-tube resonator.

On Figs 6.2A, 6.3A and 6.4A, we can see that the numerical results are consistent the theoretical predictions. The blue arrows show that the red R_4 (associated with the simple tube alone) splits into the black R_4 and R_5 (associated with the twin-tube) because the blue function is now intersected twice around its asymptote, which corresponds to the first resonant frequency of the epilaryngeal tube. The overall spectral effect is a boost around the first resonant frequency of the epilaryngeal tube that affects the neighbouring resonances and creates an extra resonance. This leads to a more appropriate definition of the Singer's Formant Cluster (SFC) as *the combination of this extra formant and a local spectral shaping of its neighbouring formants caused by the additional (epilaryngeal) tube*. This supports the results found by Titze et al. [108] where the resonance frequency of the standalone epilarynx is described as a formant attractor, which attracts the neighbouring formants.

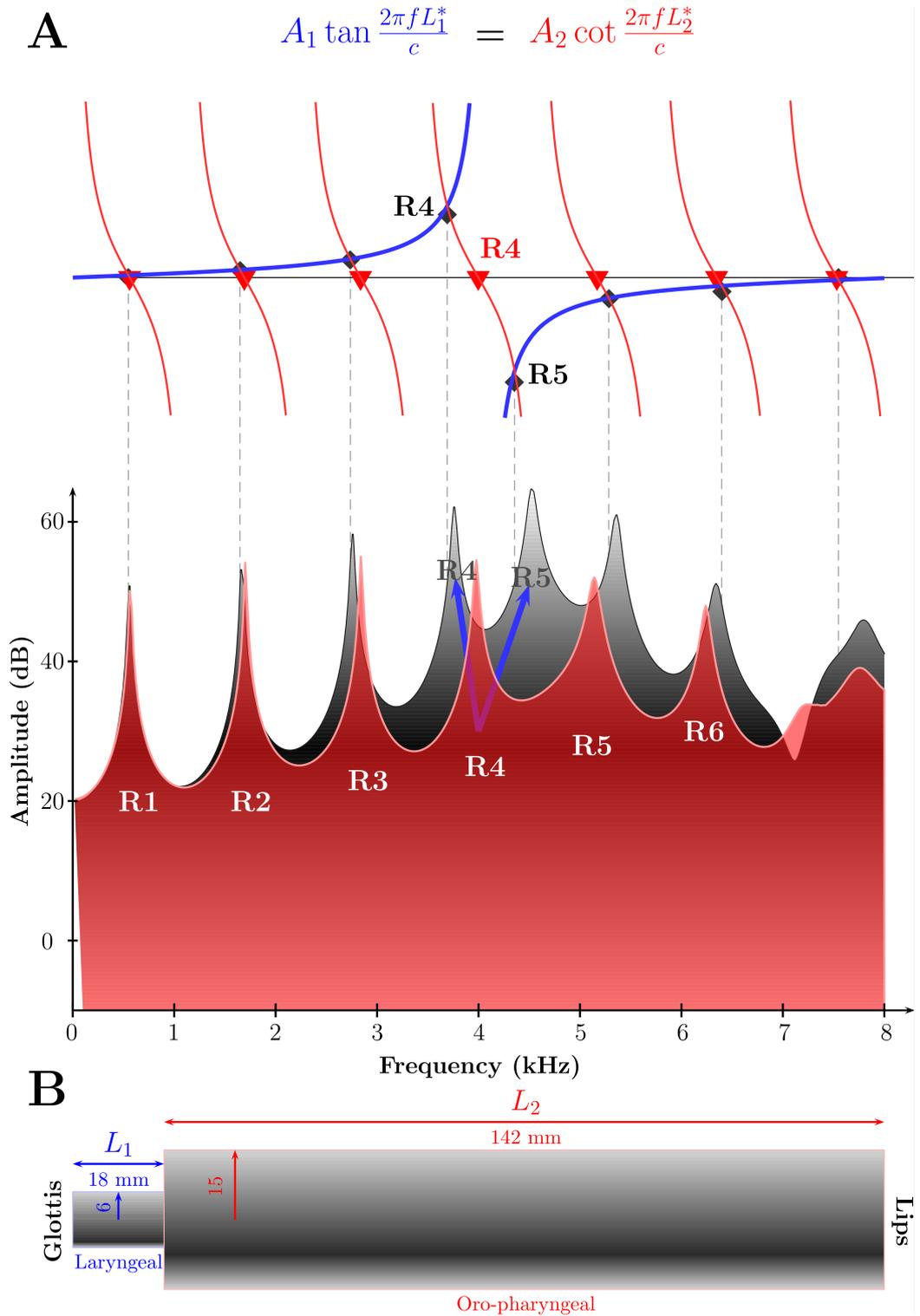


Figure 6.2: FEM numerical simulations of the resonances of one tube and a twin tube, showing the effect of appending an extra tube and generating an extra resonance. On **A**, in red, the resonances of the oro-pharyngeal tube (VTM-1): theoretical predictions (red triangles) and numerical simulation (red plot). In blue, the resonance of the epilaryngeal tube (asymptote). In black, the twin-tube (VTM-2) resonances: theoretical predictions (black diamonds) and numerical simulations (greyscale plot). **B** shows VTM-2, with the dimensions of the epilaryngeal and oro-pharyngeal tubes respectively.

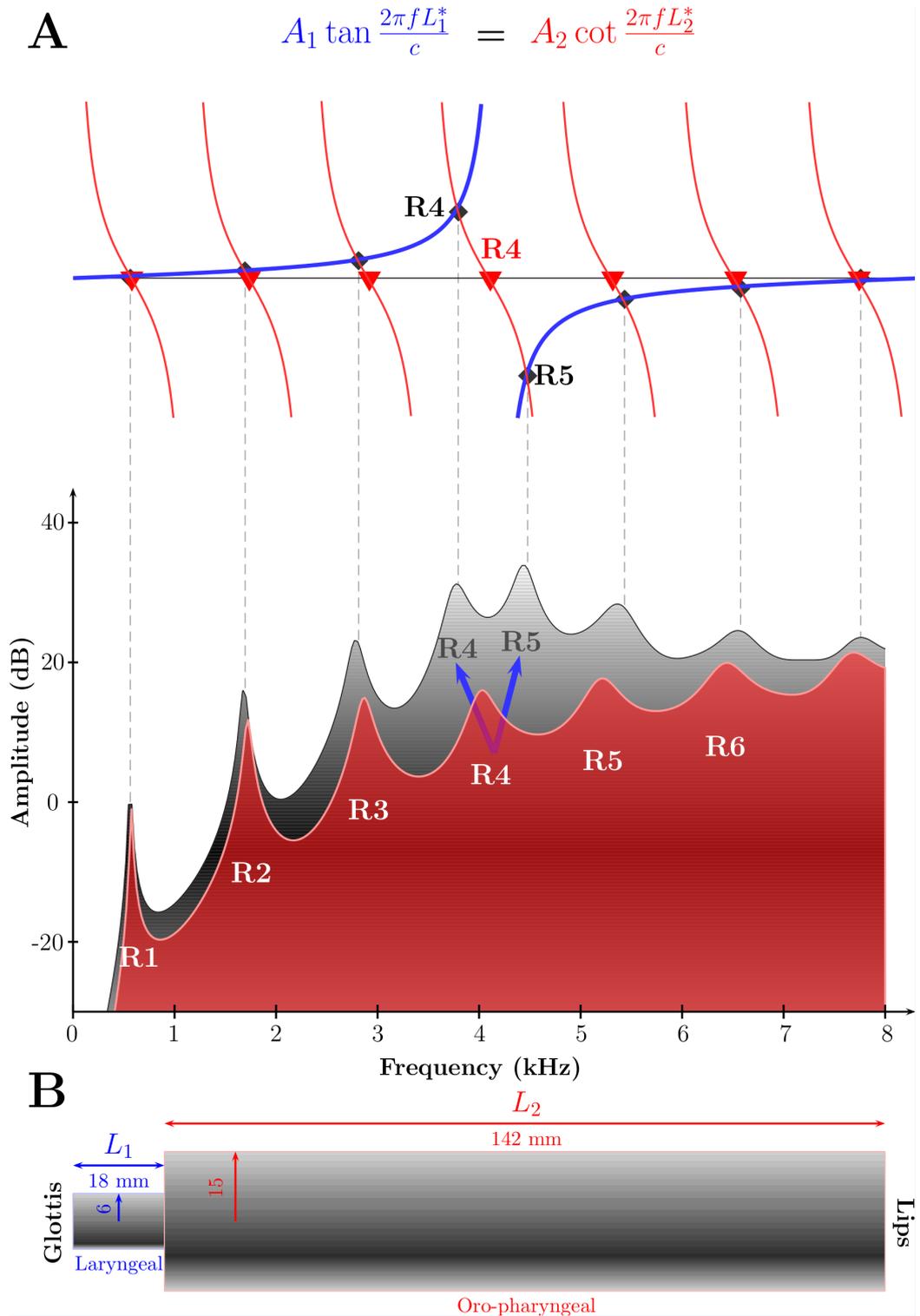


Figure 6.3: FVM numerical simulations of the resonances of one tube and a twin tube, showing the effect of appending an extra tube and generating an extra resonance. On **A**, in red, the resonances of the oro-pharyngeal tube (VTM-1): theoretical predictions (red triangles) and numerical simulation (red plot). In blue, the resonance of the epilaryngeal tube (asymptote). In black, the twin-tube (VTM-2) resonances: theoretical predictions (black diamonds) and numerical simulations (greyscale plot). **B** shows VTM-2, with the dimensions of the epilaryngeal and oro-pharyngeal tubes respectively.

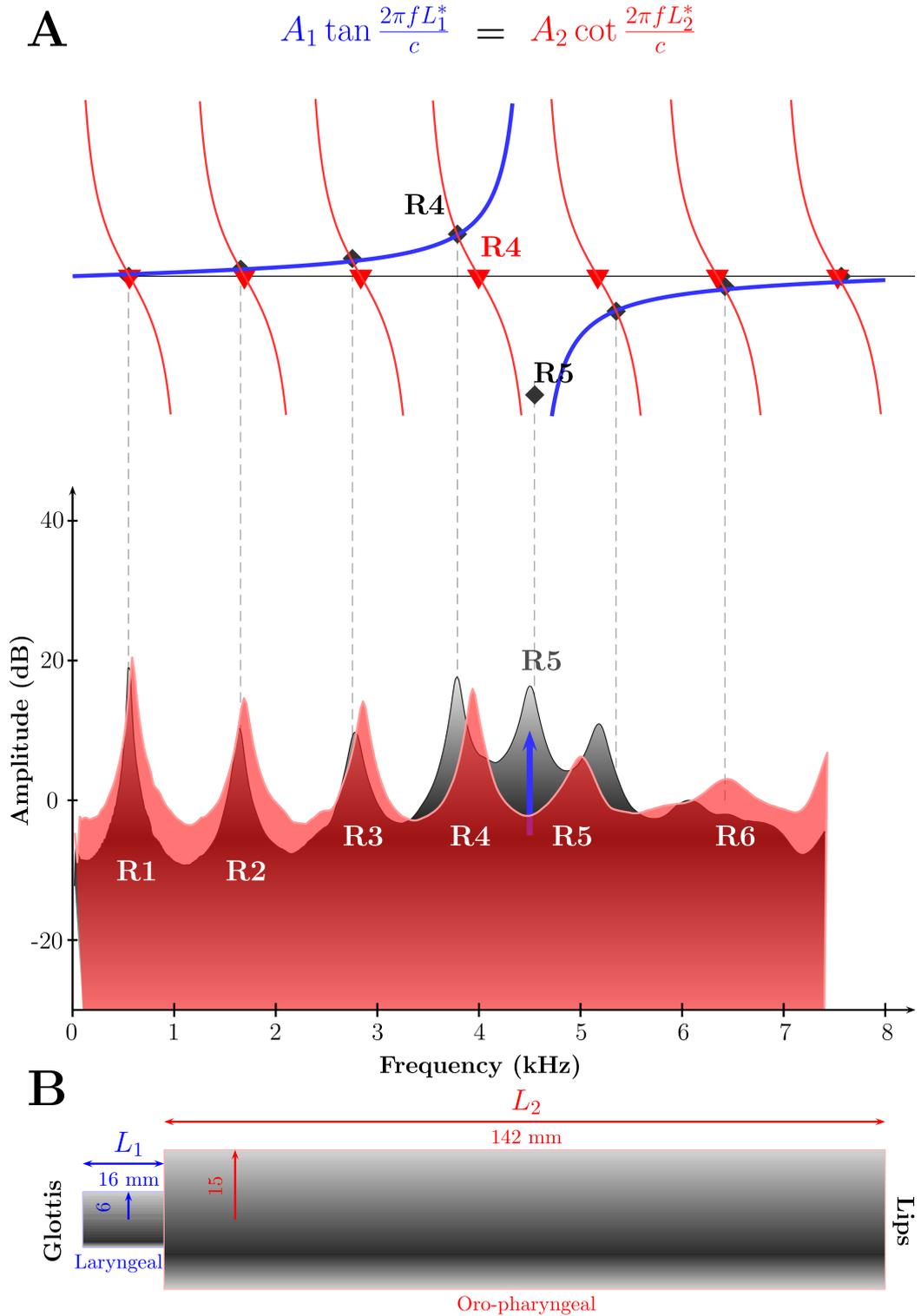


Figure 6.4: Experimental results of the resonances of one tube and a twin tube, showing the effect of appending an extra tube and generating an extra resonance. On **A**, in red, the resonances of the oropharyngeal tube (VTM-1): theoretical predictions (red triangles) and experimental results (red plot). In blue, the resonance of the epilaryngeal tube (asymptote). In black, the twin-tube (VTM-2) resonances: theoretical predictions (black diamonds) and experimental results (greyscale plot). **B** shows VTM-2, with the dimensions of the epilaryngeal and oro-pharyngeal tubes respectively.

Acoustical length

The lengths L_1^* and L_2^* used in (6.2) and (6.3) are actually the acoustical lengths of the tubes, i.e. the physical length plus the end correction which accounts for the small volume of air outside the tube vibrating along with the air inside [51]. The end correction factor is known analytically for 2 extreme cases, i.e. a cylinder with a circular flange of infinite and zero dimensions [61, 73]. The length correction for low frequencies in these two cases is $\delta_\infty = 0.8216R$ and $\delta_0 = 0.6133R$, where R is the radius of the cylinder. A fit formula for an infinite flange is given by Dalmont et al. [17] after Norris and Cheng (1989) for $kR < 3.5$:

$$\tilde{\delta}_\infty = \delta_\infty \left[1 + \frac{(0.77kR)^2}{1 + 0.77kR} \right]^{-1} \quad (6.4)$$

where $\delta_\infty = 0.8216R$, R is the radius of the inner tube and $k = \omega/c_0$ is the wavenumber.

OECC

The Open End Correction Coefficient (OECC) is the coefficient by which δ_∞ has to be multiplied to account for the finiteness of the flange (note that the end correction factor is only known analytically for 2 extreme cases, i.e. a cylinder with a circular flange of infinite and zero dimensions). Based on experimental data, Dang et al. [19] after Hall (1987) give the following empirical formula describing the relation between the OECC and the width of the flange for a low-frequency approximation,

$$\alpha_1 = 0.821 - 0.13 [(W/R) + 0.42]^{-0.54} \quad (6.5)$$

where R is the radius of the open end and W is the width of the flange.

OECC for a confined region

Ingard [51] establishes the OECC for the interior of the neck of a Helmholtz resonator:

$$\alpha_2 = 0.48(A)^{1/2} [1/r - 1.25/R] \quad (6.6)$$

where A and r are the area and the radius of the smaller-area section. R is the radius of the larger area section. This fit formula is valid as long as $r < 0.4R$, which corresponds to a 1:6 ratio, which Sundberg [101] cited as the minimum ratio between the epilaryngeal and the pharyngeal opening required to generate the SFC.

Application to a twin-tube

The resonant frequencies are given by the solution of (6.3), i.e.

$$A_1 \tan \frac{2\pi f L_1^*}{c} = A_2 \cot \frac{2\pi f L_2^*}{c} \quad (6.7)$$

L_1^* and L_2^* are the acoustical lengths of the epilaryngeal and oro-pharyngeal tubes (see Figs 6.2B, 6.3B and 6.4B), respectively:

$$\begin{cases} L_1^* = L_1 + \delta_i & \text{for the epilaryngeal tube} \\ L_2^* = L_2 - \delta_i + \delta_e & \text{for the oro-pharyngeal tube} \end{cases} \quad (6.8)$$

where δ_i and δ_e are the interior and exterior open end corrections, respectively, with

$$\begin{cases} \delta_i = \alpha_2 r \\ \delta_e = \alpha_1 R \left[1 + \frac{(0.77kR)^2}{1+0.77kR} \right]^{-1} \end{cases} \quad (6.9)$$

Note that the interior end correction δ_i is added to the epilaryngeal tube, whereas it is subtracted from the oro-pharyngeal tube.

6.2 VTM-2 : CHANGING THE DIMENSIONS OF THE EPILARYNGEAL TUBE

It was shown in subsection 6.1.2 that appending the epilaryngeal tube to the oropharyngeal tube creates an extra resonance and shapes the local spectrum around it, leading to a more appropriate definition of the SFC: *the combination of this extra formant and a local spectral shaping of its neighbouring formants caused by the additional (epilaryngeal) tube.*

Note that for the remainder of this thesis, both terms resonance and formant are used interchangeably to avoid confusion with a hypothetical Singer's Resonance Cluster (SRC) and a Singer's Formant Cluster (SFC).

In a twin-tube resonator (VTM-2), the behaviour of this SFC alters when the dimensions of the epilaryngeal tube are changed. Figs 6.5 and 6.8 show the effect of changing the length of the epilaryngeal tube whereas Figs 6.6 and 6.7 show the effect of changing its radius, for the FEM and FVM approach respectively. For all the figures, the greyscale plots represent the simulated transfer functions of VTM-2, the red plots represent the simulated transfer functions of the oropharyngeal tube (VTM-1 in this case) and the blue plots represent the first resonance frequency of the standalone epilarynx tube. The dimensions of the oro-pharyngeal tube (in red) remain the same as that of the Figs 6.2, 6.3, 6.4. Only the dimensions of the epilaryngeal tube (blue) change.

6.2.1 Changing the length of the epilaryngeal tube

Figs 6.5 and 6.8 show the effect of changing the length of the epilaryngeal tube in VTM-2 for the FEM and FVM approach respectively. The upper part of the Figs shows the first resonance frequency of the

standalone epilaryngeal tube when its length is changed. A longer length defines a lower resonant frequency (see right-hand term of (6.3)). This supports the results highlighted by Titze et al. [108] and predecessors: *"Our findings confirm the earlier results of Sundberg [101] that the epilarynx tube clusters the third, fourth, and fifth formants to generate the vocal ring (singers formant). The focal point in the spectrum is the uncoupled (free) resonance frequency of the epilarynx tube, which can be computed simply on the basis of tube length"*[108]. The lower part of the Figs shows the simulated transfer functions of VTM-2 when the epilaryngeal tube changes from 16 mm to 28 mm. The SFC (extra formant plus a local shaping of the neighbouring formants) follows the path given by the first resonance frequency of the epilaryngeal tube (upper part of the Figs), decreasing in frequency as the length of the epilarynx increases. It is very clear on Fig 6.8, where the "snowy peaks" (symptoms of a local spectral uplifting around a resonant frequency, in other words the SFC) move towards lower frequencies as the epilaryngeal length increases. Notice how the blue arrow (first resonance frequency of the epilaryngeal tube) moves from a position in between F4 and F5 (4-5 kHz) for a length of 16 mm to F3 (3 kHz) for a length of 28 mm.

6.2.2 Changing the radius of the epilaryngeal tube

Figs 6.6 and 6.7 show the effect of changing the radius of the epilaryngeal tube in VTM-2 for the FEM and FVM approach respectively. The blue arrow shows the first resonance frequency of the standalone epilaryngeal tube; as the radius of the epilaryngeal tube decreases from 9 mm to 2 mm, its frequency remains constant whereas its amplitude increases. This effect is illustrated on Fig 6.9, where the amplitude of the SFC increases as the radius decreases. Note that the change in radius effects the OECC in L_2^* , but non-significantly. Observe that both on Figs 6.6 and 6.7, F5 clusters with F4 as the radius of the epi-

laryngeal tube decreases. F_3 , to a lesser extent, is getting closer to F_4 . This is due to the fact that the attraction role of the first resonance of the epilarynx tube (the SFC is a formant attractor [108]) increases as the tube narrows. As the SFC appears and gains in amplitude, its attraction power on the neighbouring formants increases, leading to F_5 and F_3 getting closer to F_4 , "clustering" as the radius of the epilaryngeal tube decreases. This corresponds to a decrease of the SFC bandwidth.

Sundberg hypothesised that the condition needed to elicit the creation of the SFC is a ratio of at least 1:6 between the cross-sectional area of the epilarynx and the entrance of the pharynx [101]. However, on both Figs 6.6, 6.7 and 6.9, it can be seen that the SFC appears continuously. This suggests that the observation of Sundberg should rather be replaced by a gradual increase in SFC prominence.

The effects of both lengthening and narrowing the epilaryngeal tube in VTM-2 can be summarised as follows:

- Reducing the radius of the epilaryngeal tube increases the amplitude of the SFC (Figs 6.6, 6.7 and 6.9) and decreases its bandwidth accordingly, contributing to the "clustering" of the neighbouring formants of the first resonance frequency of the standalone epilaryngeal tube.
- Increasing the length of the epilaryngeal tube lowers the centre frequency of the SFC (Figs 6.5 and 6.8), along the path given by the resonant frequency of the standalone epilaryngeal tube (upper part of Figs 6.5 and 6.8, the blue curve with dots represents the asymptote of $A_1 \tan \frac{2\pi f L_1^*}{c}$ for different lengths of the epilaryngeal tube).

- The SFC centre moves from being coincident with F5 (for $L_1 = 16$ mm in Figs 6.5 and 6.8) to being between F3 and F4 for $L_1 = 28$ mm and the amplitudes of the formants around this peak are locally raised.
- The condition "*The cross-sectional area in the pharynx must be at least six times wider than that of the larynx tube opening*" [101] to obtain the SFC was not confirmed. Rather we observed a gradual increase in SFC prominence.

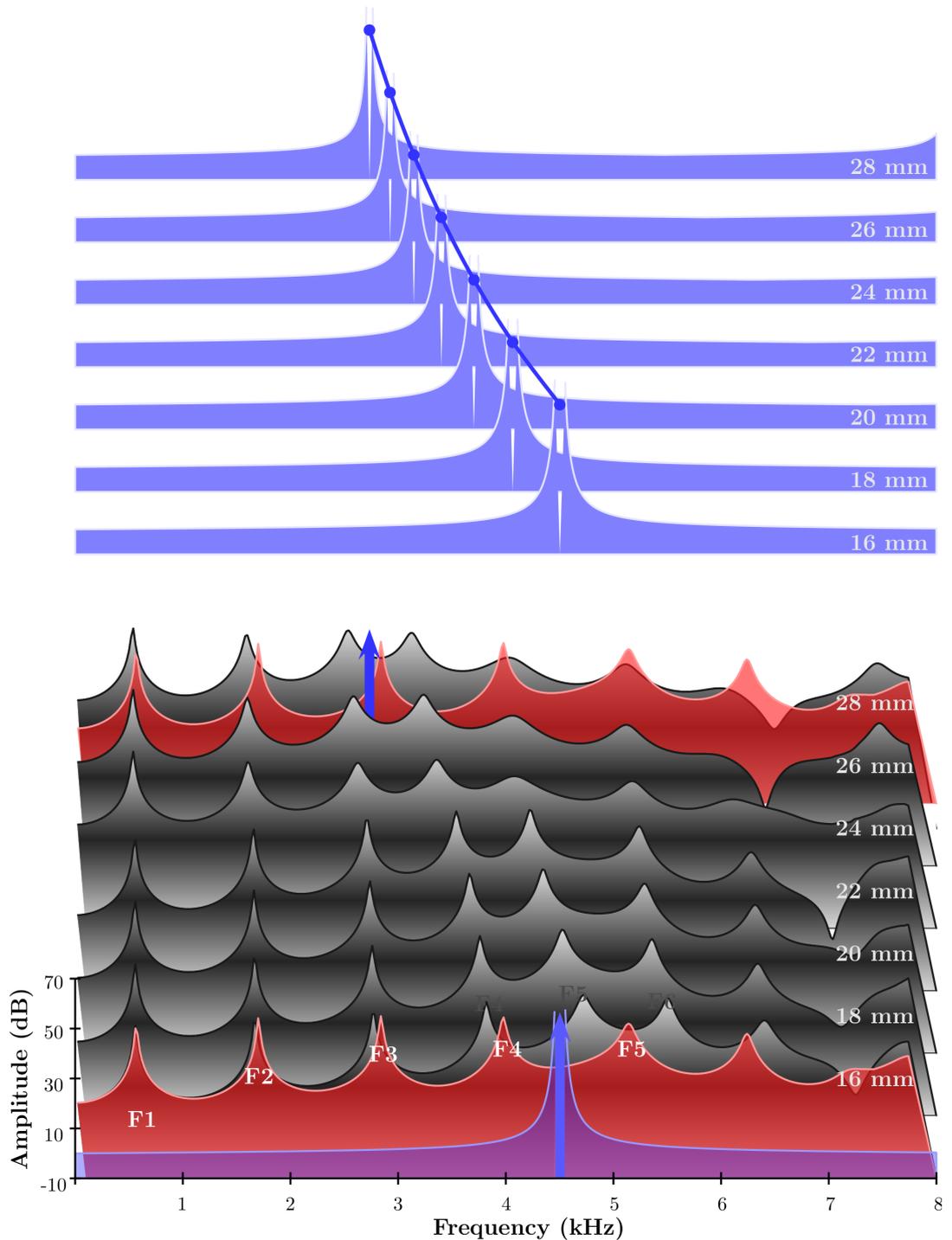


Figure 6.5: FEM simulated transfer functions when the epilaryngeal tube length of VTM-2 varies from 16 mm to 28 mm (lower part of the figure). The upper part of the figure represents the first resonance of the standalone epilaryngeal tube, which decreases in frequency as the epilaryngeal tube length varies from 16 mm to 28 mm.

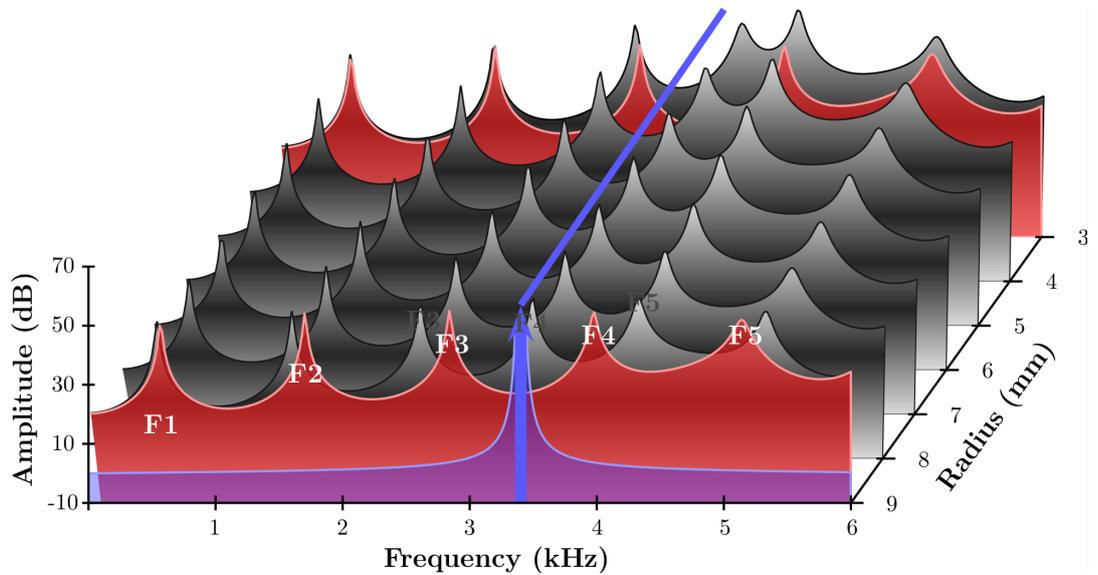


Figure 6.6: FEM simulated transfer functions when the epilaryngeal tube radius of VTM-2 varies from 9 mm to 2 mm. In blue, the evolution of the first resonance frequency of the standalone epilaryngeal tube when its radius changes from 9 mm to 2 mm.

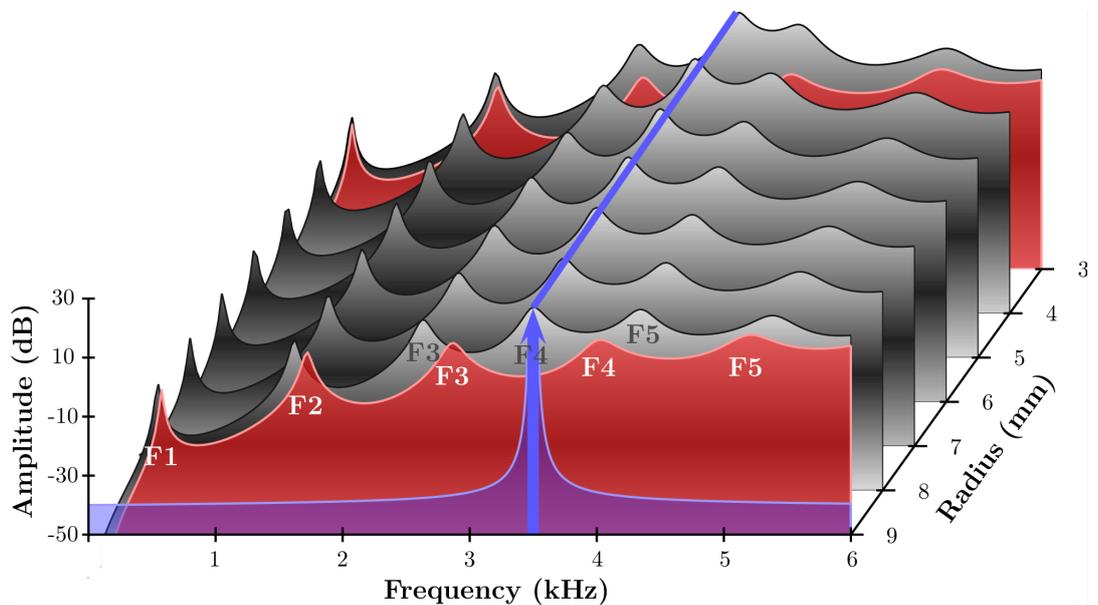


Figure 6.7: FVM simulated transfer functions when the epilaryngeal tube radius of VTM-2 varies from 9 mm to 2 mm. In blue, the evolution of the first resonance frequency of the standalone epilaryngeal tube when its radius changes from 9 mm to 2 mm.

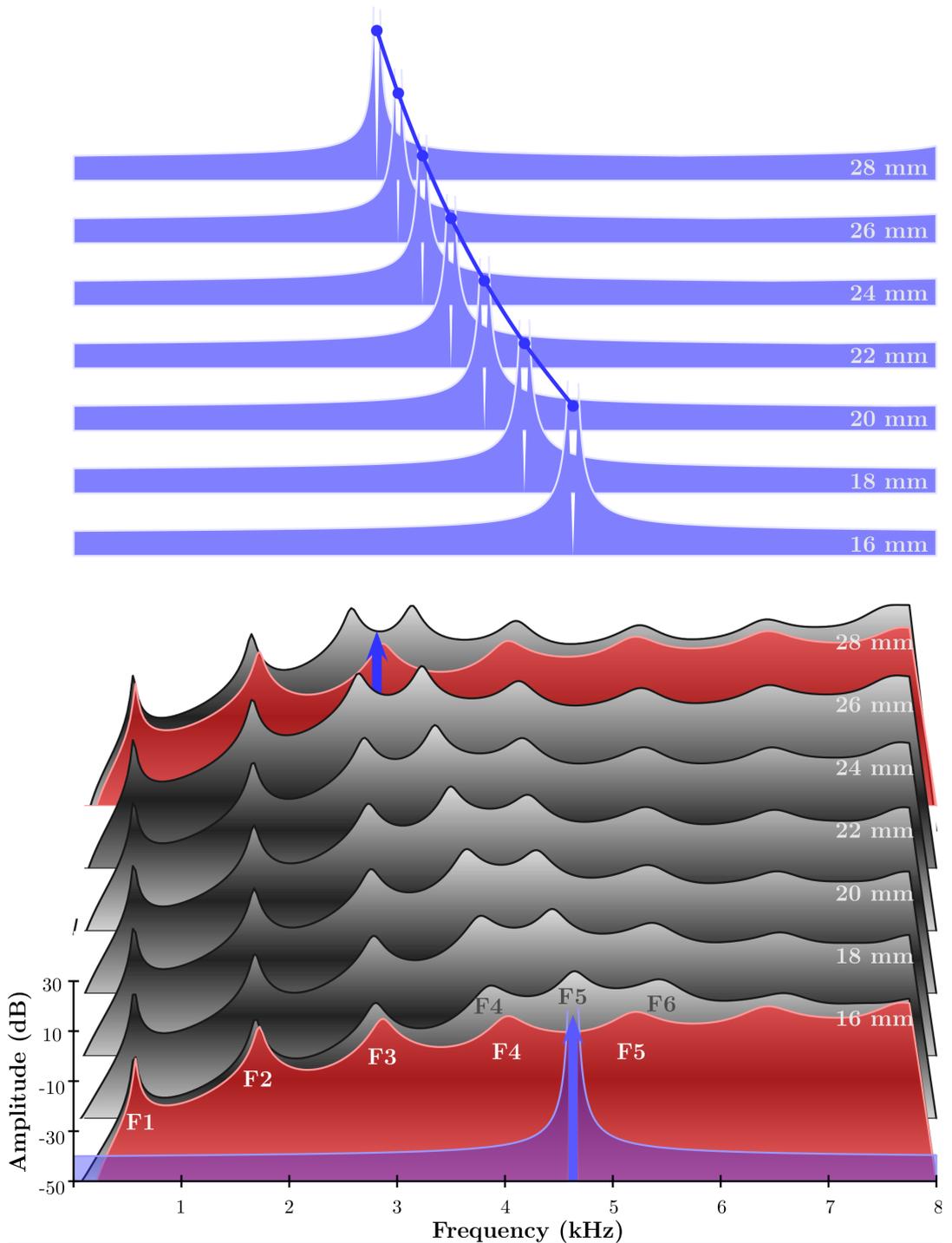


Figure 6.8: FVM simulated transfer functions when the epilaryngeal tube length of VTM-2 varies from 16 mm to 28 mm (lower part of the figure). The upper part of the figure represents the first resonance of the standalone epilaryngeal tube, which decreases in frequency as the epilaryngeal tube length varies from 16 mm to 28 mm.

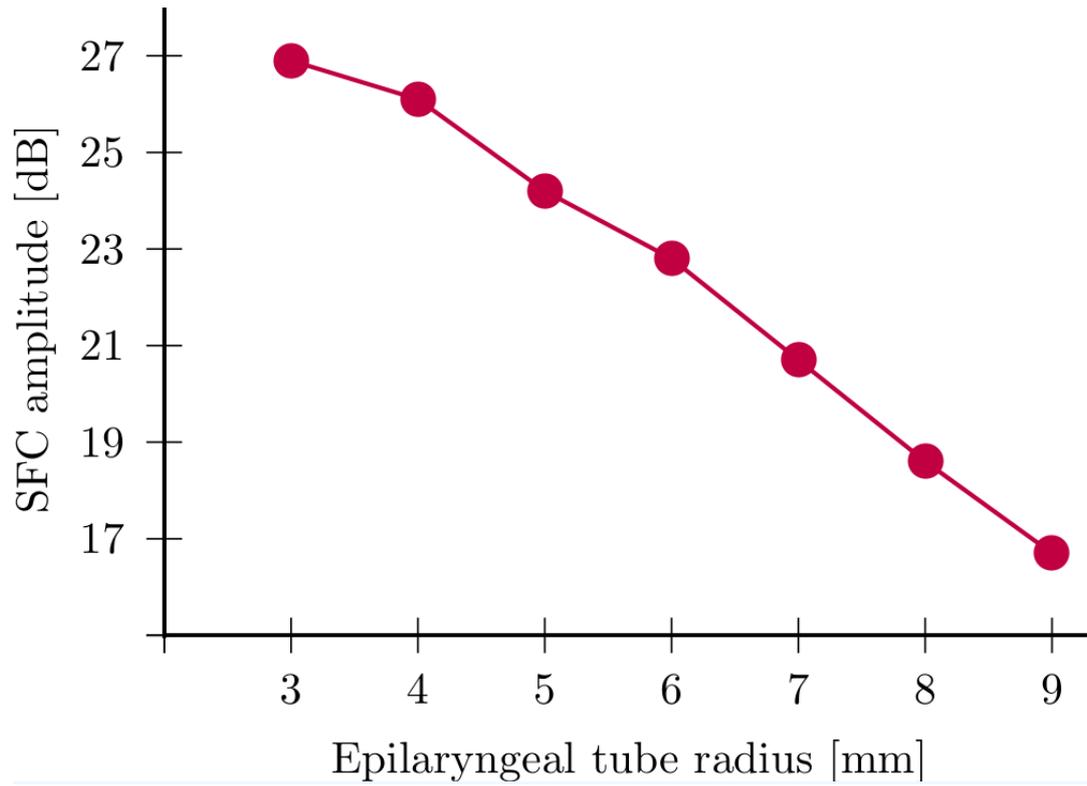


Figure 6.9: Variation of the SFC amplitude (from Fig 6.7) in relation with the epilaryngeal radius in VTM-2.

6.3 VTM-CH&K

In order to ascertain the nature of the effects observed in the previous section in a vocal tract model that is closer to reality, the dimensions of the epilaryngeal tube in the Chiba and Kajiyama's Vocal Tract Models (VTM-Ch&K) [14] are changed. These VTMs represent the cross-sectional area changes in the VT, approximated from early mid-sagittal X-Ray imaging for 5 Japanese vowels, i.e. VTM-Ch&K-/a/, VTM-Ch&K-/i/, VTM-Ch&K-/u/, VTM-Ch&K-/e/ and VTM-Ch&K-/o/.

Fig 6.10 shows the effect of changing the radius of the epilaryngeal tube of VTM-Ch&K-/i/. Fig 6.11 shows the effect of changing the length and the radius of the epilaryngeal tube in VTM-Ch&K-/a/. Figs 6.12, 6.13, 6.14, 6.15 and 6.16 show the effects of changing the length of the epilaryngeal tube in the VTM-Ch&K-/a/, VTM-Ch&K-/i/, VTM-Ch&K-/u/, VTM-Ch&K-/e/ and VTM-Ch&K-/o/.

As before, the red plot is the oro-pharyngeal tube (the VT_p, Vocal Tract proper), the grayscale plot is the full VT and the blue vertical arrow indicates the asymptote of $A_1 \tan \frac{2\pi f L_1^*}{c}$ (the first resonant frequency of the standalone epilaryngeal tube).

6.3.1 Changing the radius

Fig 6.10A shows the spectral effects on the simulated transfer function (FVM) when of VTM-Ch&K-/i/ (Fig 6.10B) when its epilaryngeal radius decreases from 10 mm to 4 mm. The blue arrows represent both the first and second resonances of the standalone epilaryngeal tube.

Note that a second SFC, as described by Titze [111] and Lee [60], is confirmed on this Fig: the metaphorical "snowy peaks" around 10-12 kHz are associated with the second resonance frequency of the standalone epilaryngeal tube. Observe that the SFC frequency centre

varies slightly (see subsection 6.2.2) whereas the changes are more visible for the second SFC frequency centre. This is due to the fact that the frequency centre (and therefore its changes) are multiplied by 3: a *quarter wavelength resonator* has resonant modes on $\frac{(2n+1)\pi}{L}$, see (6.1). Besides, the VTM-Ch&K-/i/ epilaryngeal tube has a slightly conical shape, instead of a cylinder as in VTM-1 and VTM-2. This has an effect on its first frequency resonance whose behaviour differs slightly from the theoretically known case of a cylinder.

An arbitrary amplitude threshold has been chosen to visualise (in blue) the increase in amplitude in the spectrum around the SFC. It can be seen that reducing the radius of the epilaryngeal tube from 10 mm to 4 mm increases the spectral power radiated in the frequency region surrounding the SFC and the peaks highlighted in blue are clustering around 4-5 kHz. A similar clustering is noticeable on Fig 6.11A on the right, where decreasing the radius of the epilaryngeal tube from 9 mm to 3 mm is clustering F4 and F5 around 4kHz. Note that as before, no 1:6 threshold is observed, but rather a gradual increase of the SFC prominence.

In section 6.3.2, the effect of the blue arrow (symbolising the first resonance of the epilaryngeal tube) uplifting the red plot locally (representing the simulated transfer function of the oropharyngeal tube) was defined theoretically, because the formula can be derived for two cylinders. However, in this section, the oropharyngeal tube is of a more complex shape, and the exact location of the action of the first resonance frequency cannot be theoretically determined. It can be observed that this effect is occurring close to but not precisely at the location of the first resonance frequency. This depends on the different shapes of the oropharyngeal tube, as can be seen on the Figs 6.10, 6.11, 6.12, 6.13, 6.14, 6.15 and 6.16.

6.3.2 Changing the length

Figs 6.11, 6.12, 6.13, 6.14, 6.15 and 6.16 show the effects of changing the length of the epilaryngeal tube in the VTM-Ch&K-/a/ (FVM and FEM) and VTM-Ch&K-/i/, VTM-Ch&K-/u/, VTM-Ch&K-/e/ and VTM-Ch&K-/o/ (FVM). In these Figs, **A** represents the simulated transfer function (FVM or FEM) of the VTM-Ch&K's drawn on **B**.

Increasing the length of the epilaryngeal tube of VTM-Ch&K-/a/ (Figs 6.11 and 6.12) from 16 mm to 28 mm decreases the SFC frequency centre from around 5 kHz to around 3 kHz. Changing the length of the epilaryngeal tube of VTM-Ch&K-/i/ (Fig 6.13) from 14.5 mm to 26.5 mm decreases the SFC frequency centre from around 5-6 kHz to around 3-4 kHz. Varying the length of the epilaryngeal tube of VTM-Ch&K-/u/ (Fig 6.14) from 15 mm to 27 mm decreases the SFC frequency centre from around 5 kHz to around 4 kHz. Increasing the length of the epilaryngeal tube of VTM-Ch&K-/e/ (Fig 6.15) from 14 mm to 26 mm decreases the SFC frequency centre from around 5 kHz to around 3 kHz. Changing the length of the epilaryngeal tube of VTM-Ch&K-/o/ (Fig 6.16) from 14 mm to 26 mm decreases the SFC frequency centre from around 5-6 kHz to around 3-4 kHz. Therefore, the overall trend of the SFC shift towards lower frequencies as the epilaryngeal length is increased supports the previously observed results (VTM-2, in section). This can be linked with the key difference that is observed between singers when moving from a soprano to a bass range [16]: moving from a soprano voice to a bass voice decreases the SFC frequency centre.

As can be seen on Figs 6.11, 6.12, 6.13, 6.14, 6.15 and 6.16, F_1 , F_2 (and to a lesser extent F_3) remain constant as the epilaryngeal tube length decreases: only the formants close to the "formant attractor" (name given by Titze et al. [108]) feel its local spectral influence. F_1 and F_2 being far from the attractor are essentially not affected spec-

trally. Since they are preserved, the vowel quality remains essentially the same.

As formulated before (subsection 6.3.1), the second SFC moves three times further than its equivalent (first) SFC. From the "snowy peaks", It is clear that the first resonance frequency "lifts up" its neighbouring formants to spectrally shape the SFC.

The new results can be summarised as follows :

- Decreasing the radius of the epilaryngeal tube increases the amplitude of the SFC, increasing the spectral power radiated around the SFC frequency centre (see Fig 6.10). Note that the first and second resonances of the epilaryngeal tube deviate slightly towards higher frequencies when the radius is reduced (because r varies in (6.6)).
- No threshold ratio of 1:6 is observed as reported in [101]. A gradual increase of the SFC prominence was observed instead.
- Increasing the length of the epilaryngeal tube shifts the centre of the SFC towards lower frequencies. This is a key difference that is observed between singers when moving from a soprano to a bass range [16].
- No significant variation of F_1 , F_2 (and to a lesser extent F_3) is in evidence when the epilaryngeal tube length is changed. Therefore, the vowel identification remains the same.

Fig 6.17 displays the different VTMs of Chiba & Kajiyama across the different vowels. It can be seen that below 2500 Hz, F_1 and F_2 vary according to the vowels, as expected, but above 2500 Hz, there is a common "pattern" (the region of the "snowy peaks") related to the

dimensions of the epilaryngeal tube, which is of similar dimensions across vowels.

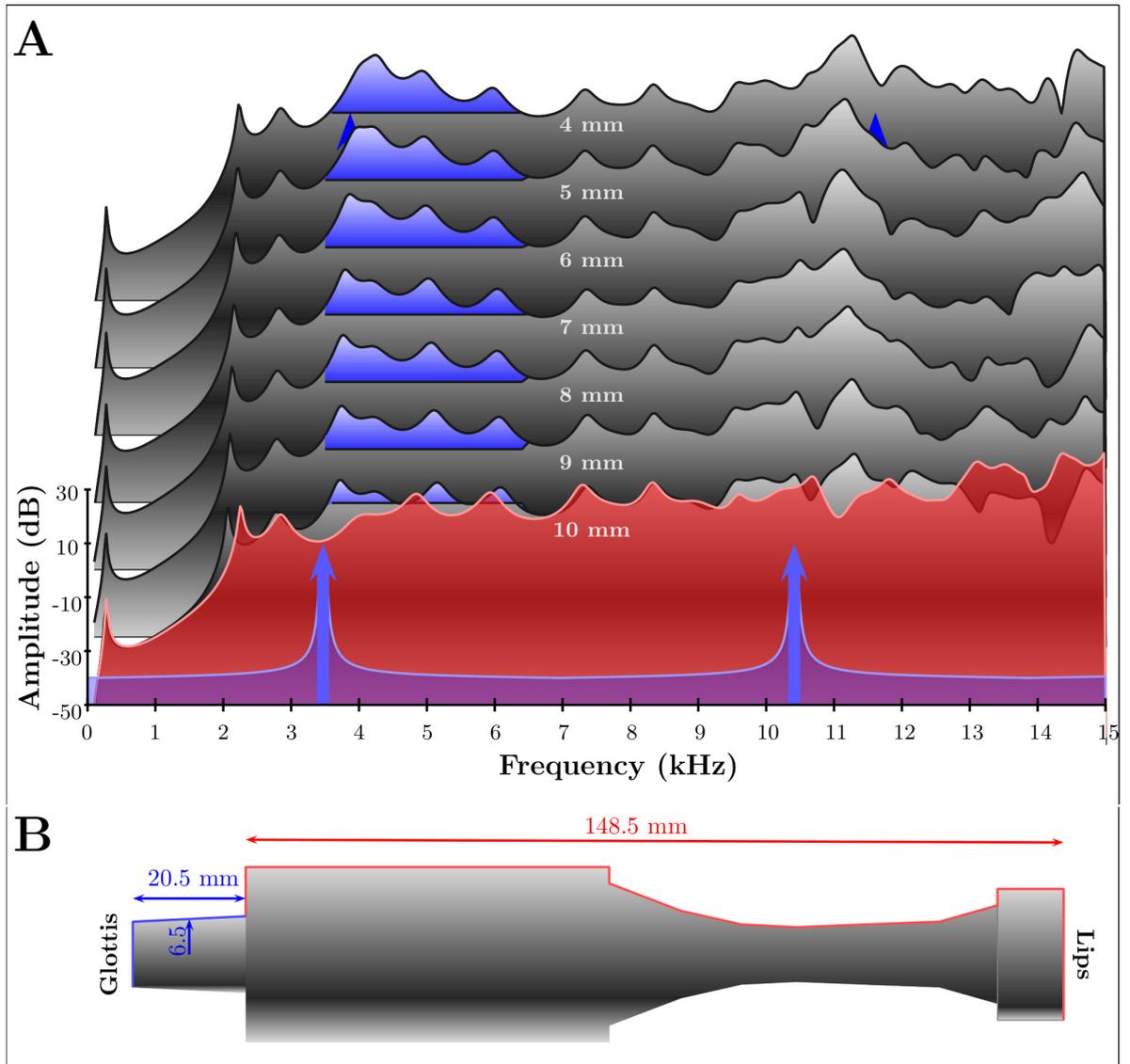


Figure 6.10: FVM simulated transfer functions (A) when changing the epilaryngeal tube radius in Chiba & Kajiyama VTM-Ch&K-/i/ B from 10 mm to 4 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube. The blue area visually aids to see the increasing radiated power of the SFC as the epilaryngeal tube radius decreases.

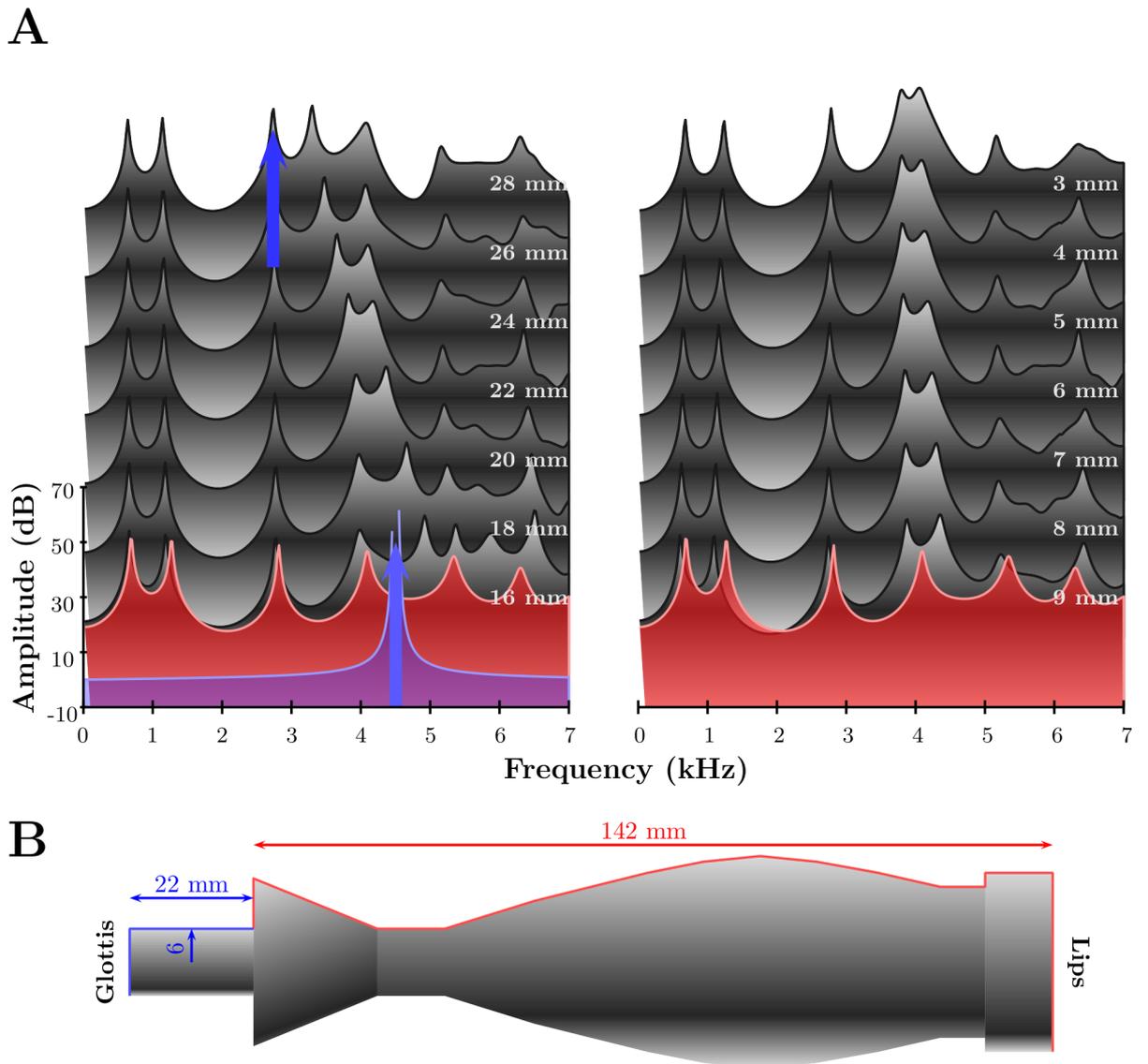


Figure 6.11: FEM simulated transfer functions (A) when changing the epilaryngeal tube length and radius in Chiba & Kajiyama VTM-Ch&K-/a/ B from 16 mm to 28 mm and from 9 mm to 3 mm respectively. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube.

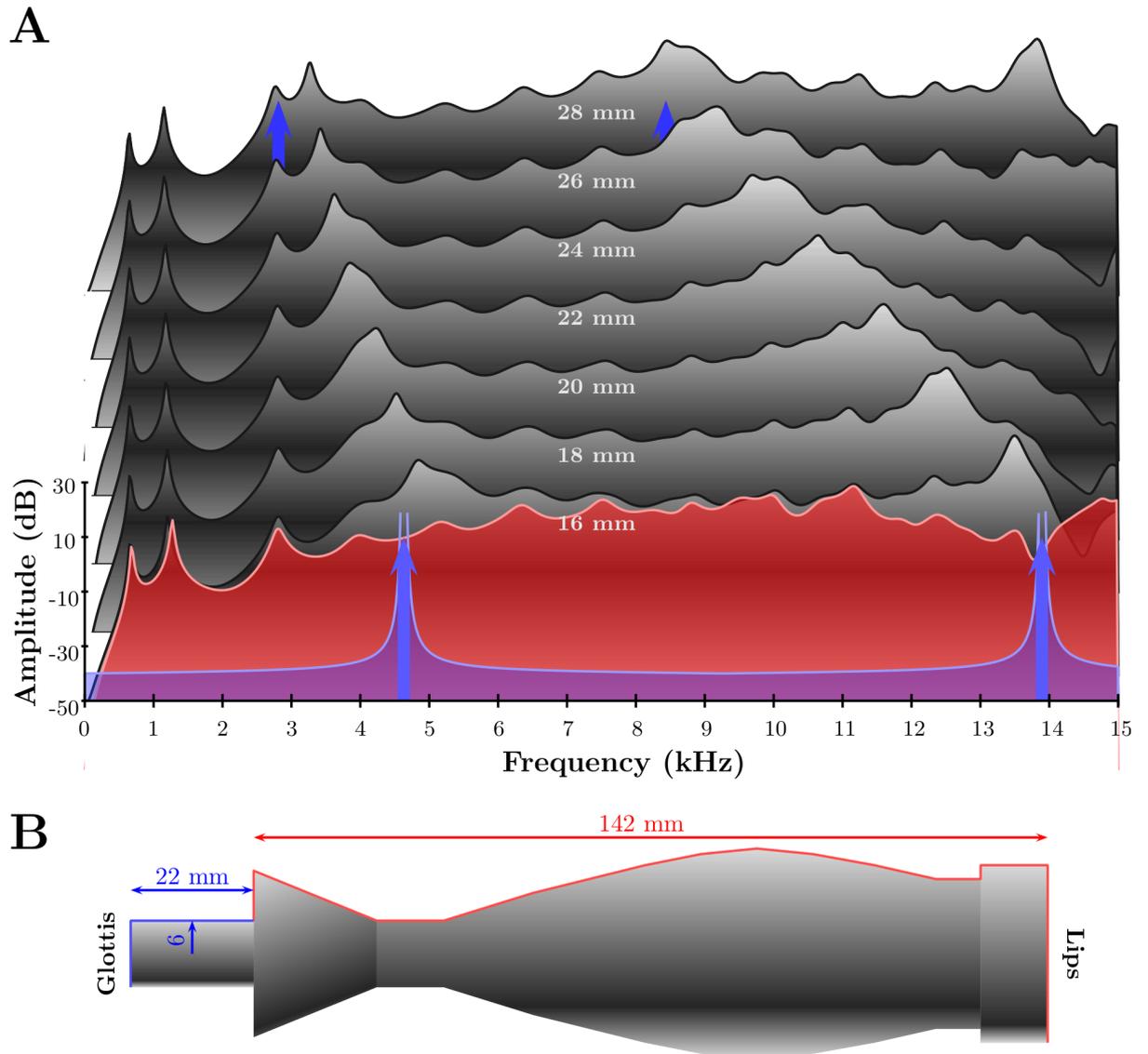


Figure 6.12: FVM simulated transfer functions (A) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/a/ B from 16 mm to 18 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube.

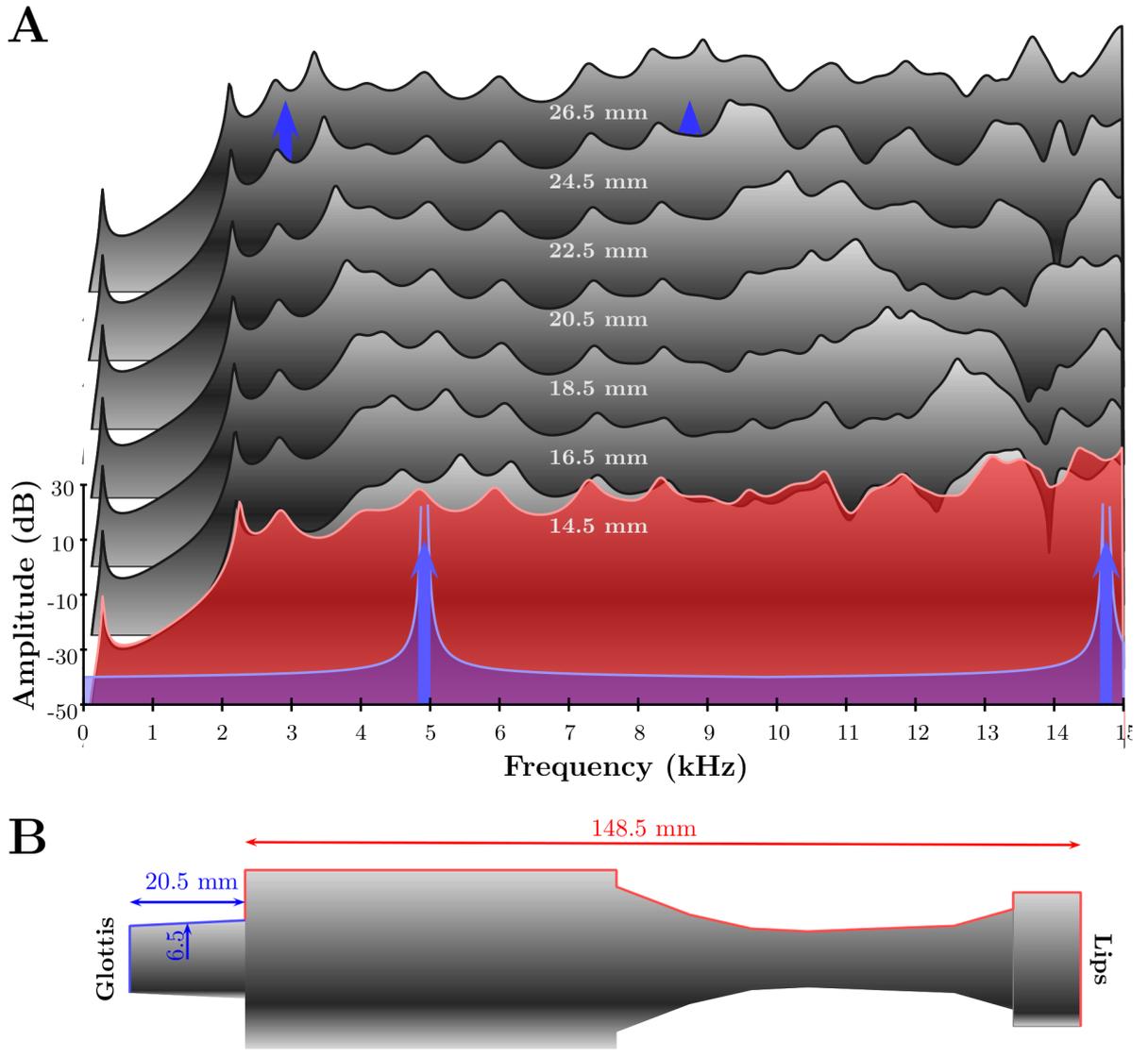


Figure 6.13: FVM simulated transfer functions (A) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/i/ B from 14.5 mm to 26.5 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube.

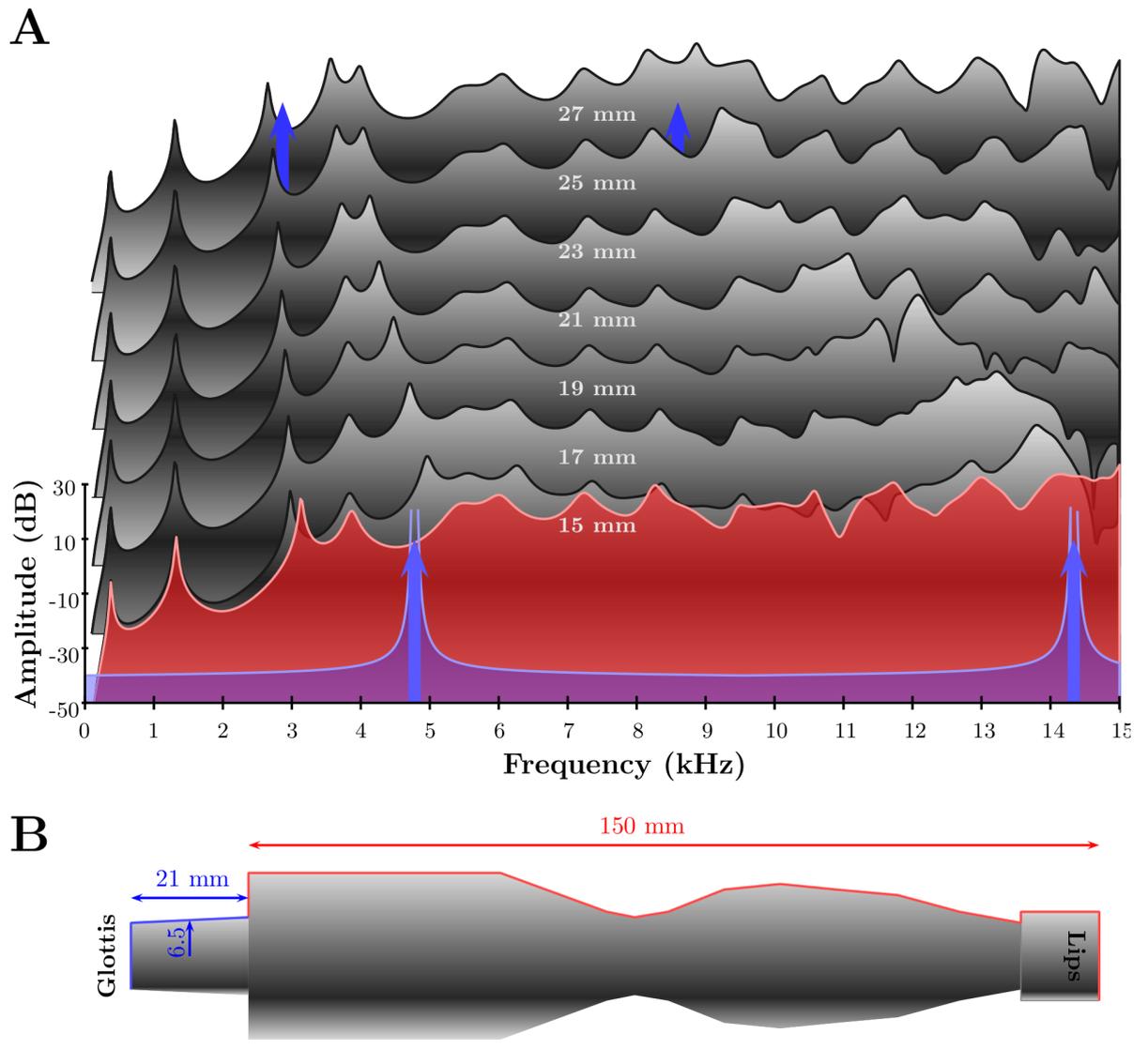


Figure 6.14: FVM simulated transfer functions (A) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/a/ B from 15 mm to 27 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube.

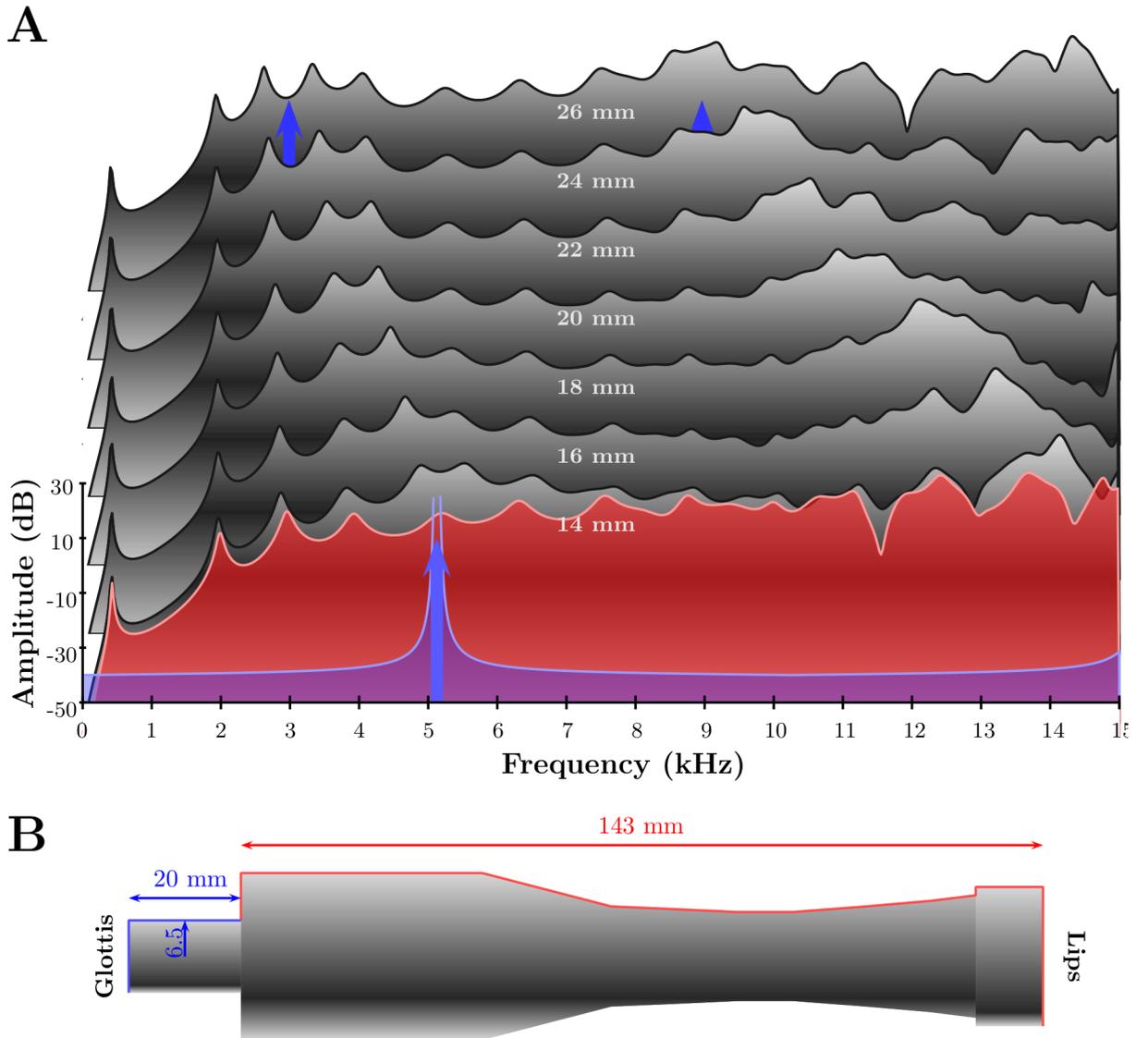


Figure 6.15: FVM simulated transfer functions (A) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/e/ **B** from 14 mm to 26 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube.

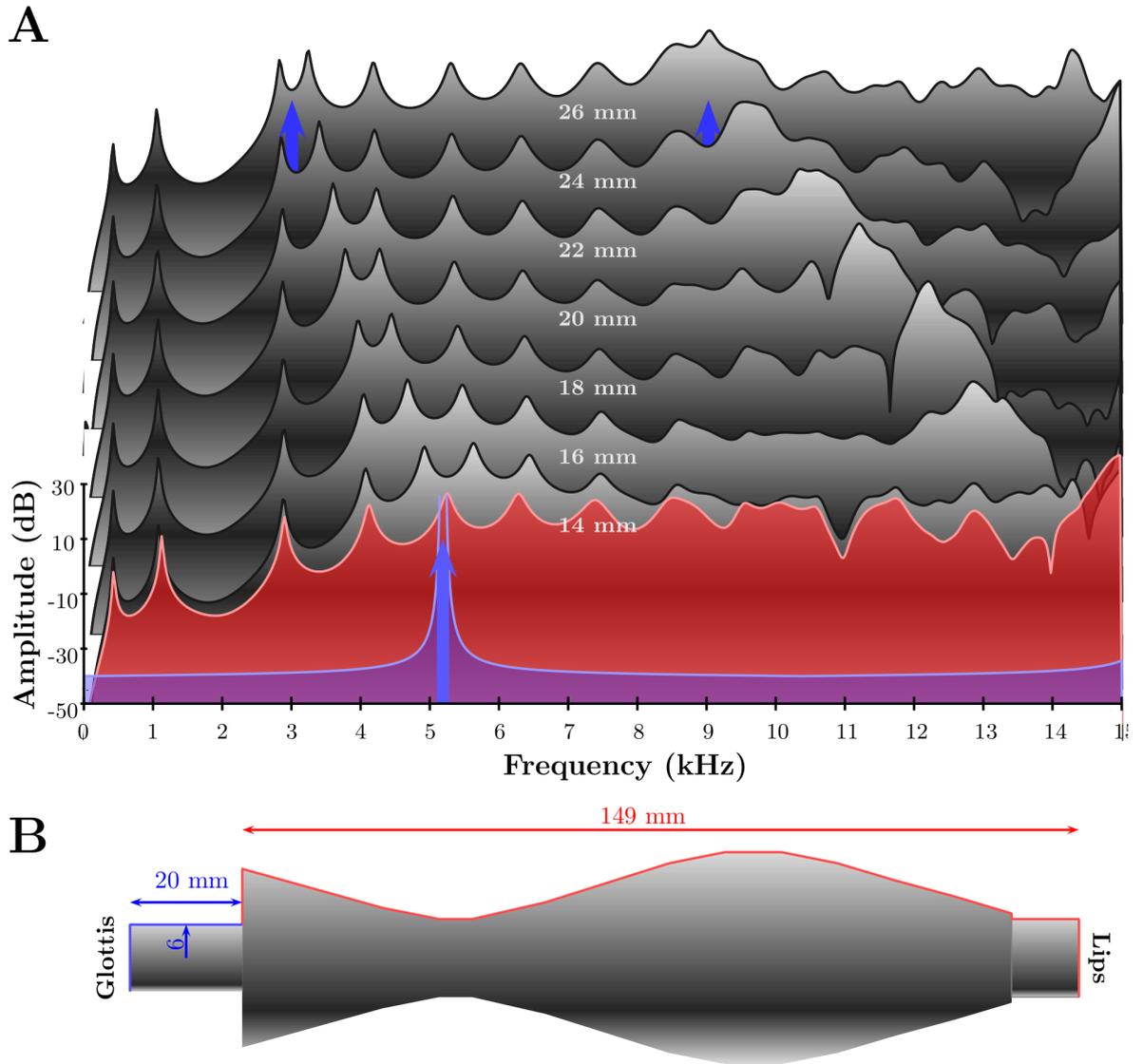


Figure 6.16: FVM simulated transfer functions (A) when changing the epilaryngeal tube length in Chiba & Kajiyama VTM-Ch&K-/o/ B from 14 mm to 26 mm. In red, the oropharyngeal transfer function. The blue arrows indicate the resonances of the standalone epilaryngeal tube.

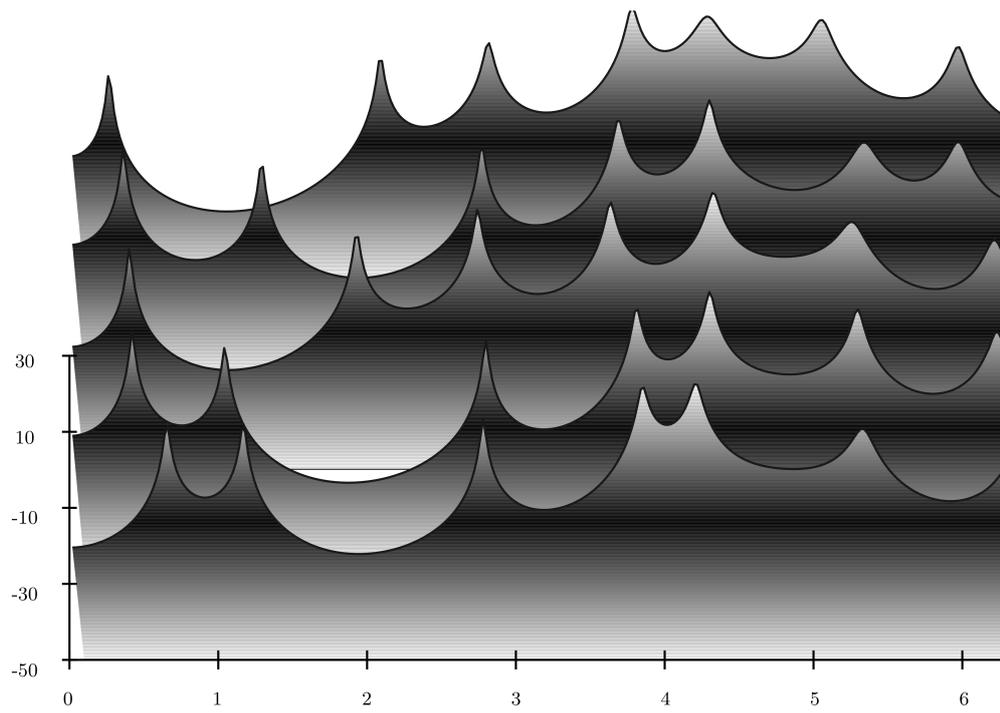


Figure 6.17: FEM simulated transfer functions for VTM-Ch&K. From top to bottom, VTM-Ch&K-/i/, VTM-Ch&K-/u/, VTM-Ch&K-/e/, VTM-Ch&K-/o/, VTM-Ch&K-/a/.

6.4 VTM-MRI

This section contains the numerical simulations (FEM) of the transfer functions of the VTM-MRI of 6 professional singers (spreading across voice categories). The corpus is composed of 1 Soprano, 2 Mezzo-Sopranos, 1 Tenor, 1 Bari-Tenor and 1 Bass-Baritone (for more details, see part ii).

6.4.1 Comparison across voice categories

Fig 6.18 displays the simulated transfer functions (FEM) of the different singers, from the Bass-Baritone (Barnaby) to the Soprano (Sophy) singing on /hard/. Table 6.1 shows the pitch sung, along with the five first resonance frequencies, the average of the third, fourth and fifth resonant frequencies and their corresponding statistical average and standard deviation.

A first observation shows that some singers diverge from the expected values of R1 and R2 vowels for /hard/ ($\mu(R1) = 612$ Hz and $\mu(R2) = 1200$ Hz respectively): Sophy for R1 and R2, Maristela for R2, Bartholomew for R1. But the other singers have a fairly constant combination of R1-R2. Sophy is probably using the resonance tuning strategy, tuning R1 (866 Hz) slightly above f_0 (784 Hz) and R2 (1519 Hz) slightly under $2f_0$ (1568 Hz). Maristela seems to use as well the tuning resonance strategy to tune R2 (1322 Hz) to $5f_0$ (1310 Hz). Regarding Bartholomew, a possible explanation as to the lower value of R1 (441 Hz for $\mu(R1) = 612$ Hz) is the fact that his segmented VT has been cleared from a velar opening leading to a resonance cavity. Since the canals and sinuses linking this cavity to the nares are very thin and blurry on the MR image, it was chosen to remove this velar connection. It is thought that removing the acoustical side branch might

alter the first resonance. Future study will investigate the effect of an open velum on the first resonance.

Note that the statistical distribution the five first resonances shows a different spreading across the vowel-type resonances (R_1 , R_2) and the timbre-type resonances (R_3 , R_4 and R_5) in Table 6.1. The vowel-type resonances spread are statistically less spread across the spectrum ($\sigma = 139$ Hz and $\sigma = 187$ Hz respectively than the timbre-type resonances ($\sigma = 364$ Hz, $\sigma = 540$ Hz and $\sigma = 613$ Hz respectively). This supports the fact that F_1 , F_2 are associated with vowel quality [78, 72, 102, 47, 58] and show small inter-individual variations [57], whereas F_3 , F_4 , F_5 are associated with voice quality [99, 119, 58, 32, 102] and show large inter-individual variations [57]. Regarding higher resonances, the last column of Table 6.1 shows that the group R_3 - R_4 - R_5 shifts towards higher frequencies when the voice type is higher: in order of lower voice to higher voice, $\mu(R_3, R_4, R_5) = 2746$ Hz, 2777 Hz, 3044 Hz, 3613 Hz, 3552 Hz, 3923 Hz respectively, with $\mu(\mu(R_3, R_4, R_5)) = 3276$ Hz and $\sigma(\mu(R_3, R_4, R_5)) = 488$ Hz. Note that $\mu(R_3, R_4, R_5)$ increases of 43% from for Barnaby to Sophy. This supports the results previously introduced with VTM-2 and VTM-Ch&K (see section 6.3): as the epilarynx length increases, the SFC frequency centre decreases. This can be seen on Fig 6.18: the "snowy peaks" corresponding to the SFC are migrating towards lower frequencies as the voice category goes from the highest voices (soprano) to the lowest voices (bass). The length of the epilarynx for each singer will be examined in the next section.

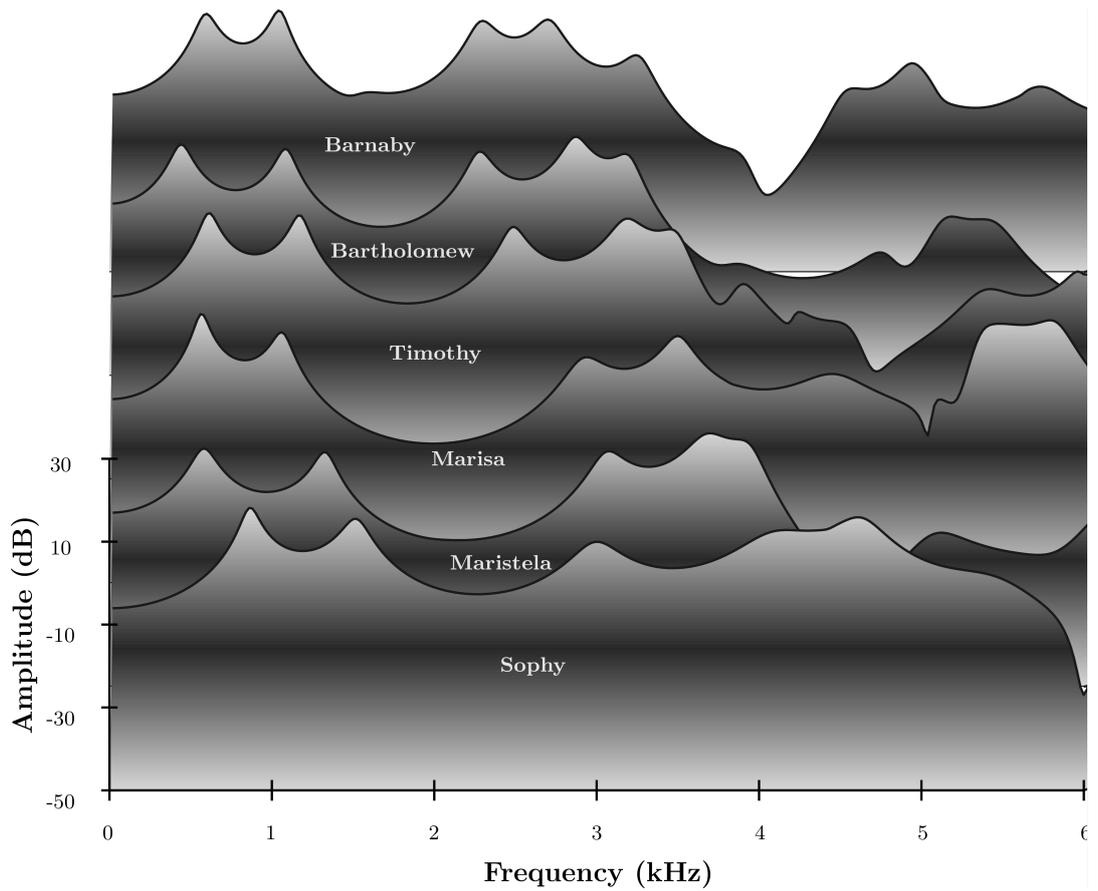


Figure 6.18: FEM transfer functions of different voice types phonating on /hard/. From top to bottom, Barnaby, Bartholomew, Timothy, Marisa, Maristela, Sophy. Details of the phonation and resonance data are given in Table 6.1.

Resonance frequencies across voice categories

	Pitch	f_0	R1	R2	R3	R4	R5	$\mu(R_3, R_4, R_5)$
Sophy	G5	784	866	1519	3005	4153	4611	3923
Maristela	C4	262	579	1322	3075	3699	3883	3552
Marisa	A4	440	569	1060	2919	3489	4432	3613
Timothy	F4	349	617	1169	2486	3192	3456	3044
Bartholomew	F#3	185	441	1082	2279	2874	3178	2777
Barnaby	G#2	104	599	1045	2298	2699	3241	2746
μ			612	1200	2677	3351	3800	3276
σ			139	187	364	540	613	488

Table 6.1: Sung pitch, corresponding f_0 , five first resonance frequencies (in Hz) across voice categories for VTM-MRI-/hard/, average of R3, R4 and R5 and their statistical distribution (average μ and standard deviation σ).

6.4.2 Dimensions of the Vocal Tract

The dimensions of the Vocal Tract across voice categories have been listed in Table 6.2. EV is the Epilaryngeal Volume (cm^3), ECS the epilaryngeal Cross-Section (cm^2), ER the equivalent Epilaryngeal Radius (mm), VTL the Vocal Tract Length (mm), OL the Oral Length (mm), PL the Pharyngeal Length (mm), EL the Epilaryngeal Length (mm), PW the Pharyngeal Width (mm), VTV the Vocal Tract Volume (cm^3), PV the piriform fossae volume (cm^3), % the ratio of PV/VTV expressed in %, SFC (pred) (in Hz) the SFC frequency centre predicted from the dimensions of the hypopharynx (see subsection 6.4.2.1).

ER, the equivalent Epilaryngeal radius is calculated as follows:

$$ER = \sqrt{\frac{ECS}{\pi}} \quad (6.10)$$

ER therefore represents the radius that the epilaryngeal opening would have at the pharyngeal opening if it were circular.

Dimensions of the Vocal Tract and prediction of the SFC center												
	EV	ECS	ER	VTL	OL	PL	EL	PW	VTV	PV	%	SFC (pred)
Sophy	1.09	1.18	6.13	158.63	83.62	58.71	16.30	24.39	82.54	1.7	2.06	4,002.19
Maristela	0.81	1.29	6.41	152.01	81.18	52.11	18.72	36.00	44.58	3.47	7.78	3,529.05
Marisa	0.77	6.36	14.23	190.96	89.41	83.62	17.93	30.28	111.82	1.8	1.61	3,529.33
Timothy	1.79	2.09	8.15	178.16	89.93	65.78	22.45	28.90	67.26	2.51	3.73	2,971.09
Bartholomew	2.58	1.68	7.31	187.82	99.35	64.01	24.46	36.50	37.32	1.7	4.56	2,789.19
Barnaby	1.36	1.16	6.07	202.08	88.04	87.47	26.57	33.43	80.49	1.69	2.1	2,671.22
μ	1.40	2.29	8.05	178.28	88.59	68.62	21.07	31.58	70.67	2.15	3.64	3249
σ	0.69	2.02	3.13	19.47	6.30	14.00	4.04	4.64	27.33	0.723	2.32	519
σ in % of μ	49.18	88.26	38.88	10.92	7.11	20.41	19.19	14.7	38.68	33.69	63.85	15.98

Table 6.2: Vocal Tract dimensions versus SFC prediction. Epilaryngeal Volume (cm^3), Epilaryngeal Cross-Section (cm^2), equivalent Epilaryngeal Radius (mm), Vocal Tract Length (mm), Oral Length (mm), Pharyngeal Length (mm), Epilaryngeal Length (mm), Pharyngeal Width (mm), Vocal Tract Volume (cm^3), Piriform fossae Volume (cm^3), % the ratio of PV/VTV (in %), SFC (pred) (in Hz) the SFC frequency centre predicted from the dimensions of the hypopharynx (see subsection 6.4.2.1)

The largest variations of the VT dimensions are observed in the epilarynx with a standard variation σ (reported in percentage of the average μ) raising to 49 % in volume across individuals and 88 % in cross-section. The third column corresponds to a fictive epilaryngeal dimension and cannot be compared as such across singers. Both the volumetric and the cross-sectional dimensions of the epilarynx vary on average by 69 % around their respective average dimension across singers. In comparison, the average of the length dimensions variation around their average value is 15 %, so more than a fourfold difference between the relative variations of the Vocal Tract length-related dimensions to the epilaryngeal dimensions across singers. The volumetric variations of the vocal tract, the piriform fossae and the epilarynx across singers are of 39 %, 34% and 49% respectively. The hypopharyngeal volumetric variation (piriform fossae + epilarynx) across singers is 41 %. This supports the large inter-individual variations of the hypopharynx, as reported by Kitamura et al. [57]

6.4.2.1 Prediction of the SFC center according to VT dimensions

The equations (6.8, 6.9, 6.6) give the first resonance frequency of the epilaryngeal tube in the twin-tube model, by:

$$f = \frac{c}{4 \cdot \left(L + 0.48\sqrt{A} \left(\frac{1}{r} - \frac{1}{1.25R} \right) \cdot r \right)} \quad (6.11)$$

where c is the speed of sound, L the length of the epilaryngeal tube, A the cross-section of the epilaryngeal tube, r the radius of the epilaryngeal tube and R the radius of the oropharyngeal tube.

The geometry of an actual VT is much more complicated. Inspired by (6.11), the SFC centre is suggested to be predicted based on the dimensions of the epilaryngeal tube and the pharyngeal opening by replacing L by EL , A by ECS , r by ER and R by $PW/2$, leading to the formula:

$$f = \frac{c}{4 \cdot \left(L + 0.48 \cdot \left(\sqrt{A} - \frac{2.5A\sqrt{\pi}}{\pi PW} \right) \right)} \quad (6.12)$$

Note that EL is measured from the mid-point of the glottis to the centre of the cross-section of the epilaryngeal tube at the pharyngeal opening and PW measures the width of the pharynx at the pharyngeal opening, as shown in Fig 6.20.

The predicted values of SFC frequency centre following the suggested formula (6.12) are given as the last column of Table 6.2. The inverse correlation between the epilarynx length and the SFC frequency centre predicted from 6.12 is shown on Fig 6.19: the lower the voice category, the longer the epilarynx, the lower the SFC frequency centre. This supports the findings of Dmitriev et al. [25] (Fig 2.14): they found that the dimensions of the VT were related to the production of singing formant and to the voice classification.

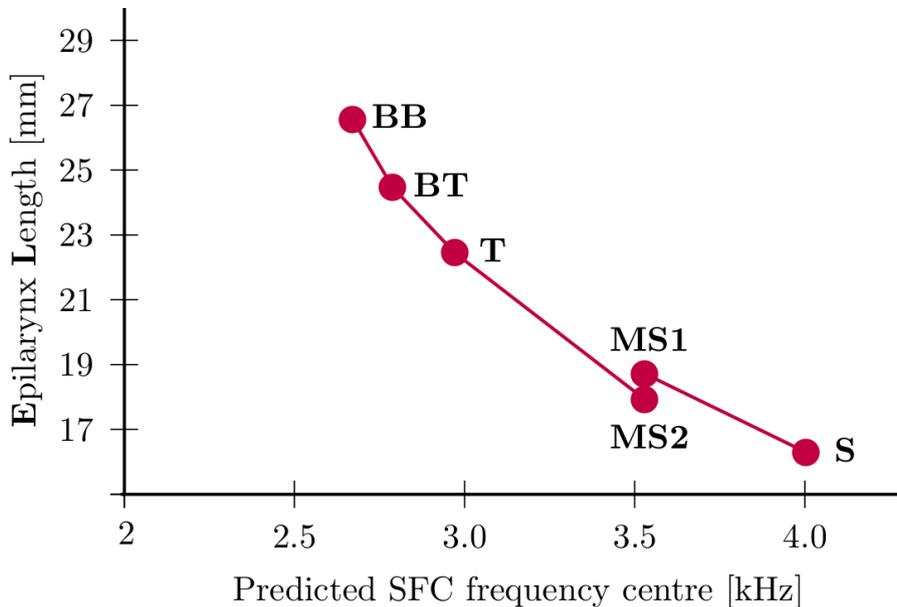
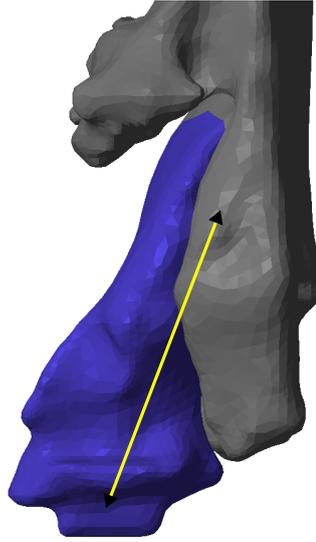


Figure 6.19: Variation of the SFC frequency centre predicted by 6.12 in relation with the Epilarynx Length across singers: Soprano Sophy, Mezzo-Soprano 1 Maristela, Mezzo-Soprano 2 Marisa, Tenor Timothy, Bari-Tenor Bartholomew and Bass-Baritone Barnaby.

Epilaryngeal length



Pharyngeal width

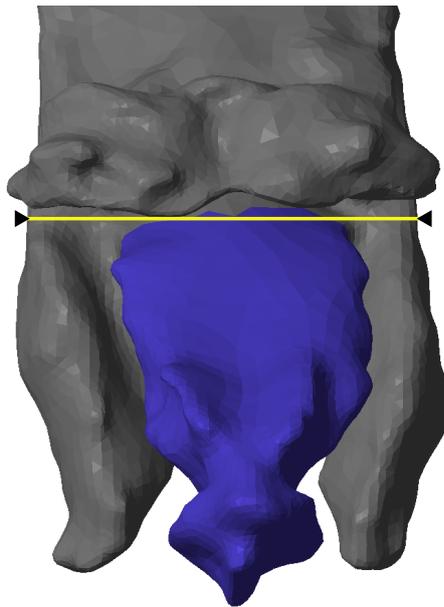


Figure 6.20: Epilaryngeal length (EL) and Pharyngeal width (PW) in the suggested formula 6.12 to predict the SFC centre.

6.4.3 Effect of the epilarynx

This subsection examines the spectral impact of the epilarynx when it is appended to the Vocal Tract proper (terminology introduced by [106] standing for the Vocal Tract whose epilarynx has been removed) and when it is lengthened downwards.

6.4.3.1 *Vocal Tract proper vs Vocal Tract*

The Figs 6.21, 6.22, 6.23, 6.24, 6.25, 6.27, and 6.27, 6.28, 6.29, 6.30, 6.31 show the difference in the transfer function between the Vocal Tract (VT) and the Vocal Tract proper (VTp) of different voice types (Sophy, Marisa, Maristela, Timothy, Bartholomew and Barnaby) and Barnabas singing on different vowels (/hard/, /port/, /stern/, /food/ and /neap/ respectively). The transfer function of the VT is plotted in grayscale whereas the one of the VTp is plotted in red atop. The blue arrow indicates the SFC centre predicted by the suggested formula (6.12), based on the hypopharynx dimensions.

On Fig 6.21 (Sophy), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 4002 Hz, near F₄ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₄ splits into F₄ and F₅ (around 4-5 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. Note that the data of Marisa were of bad quality (see section ii): this is the reason why Fig 6.22 does not provide the expected results as the other singers did. On Fig 6.23 (Maristela), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 3529 Hz, near F₄ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₄ splits into F₄ and F₅ (around 4 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. On Fig 6.24 (Timothy), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2971 Hz between F₃ and F₄ on the VTp transfer function. Appending the epilarynx results in the creation of

an extra formant: F₄ splits into F₄ and F₅ (above 3 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. On Fig 6.25 (Bartholomew), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2789 Hz between F₃ and F₄ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₄ splits into F₄ and F₅ (around 3 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. On Fig 6.27 (Barnaby), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2671 Hz near F₃ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₃ splits into F₃ and F₄ (under 3 kHz) and the spectrum amplitude is locally lifted up around the blue arrow.

On Fig 6.28 (Barnaby singing on /port/), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2707 Hz near F₄ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₄ splits into F₃ and F₄ (slightly under 3 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. On Fig 6.29 (Barnaby singing on /stern/), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2769 Hz between F₃ and F₄ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₄ arises between F₃ and F₄ of the VTp (slightly under 3 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. On Fig 6.30 (Barnaby singing on /food/), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2796 Hz near F₄ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₄ splits into F₄ and F₅ (around 3 kHz) and the spectrum amplitude is locally lifted up around the blue arrow. On Fig 6.31 (Barnaby singing on /neap/), it can be seen that the predicted SFC frequency centre (blue arrow) arises at 2508 Hz near F₃ on the VTp transfer function. Appending the epilarynx results in the creation of an extra formant: F₃ splits into F₃ and F₄ (under 3

kHz) and the spectrum amplitude is locally lifted up around the blue arrow.

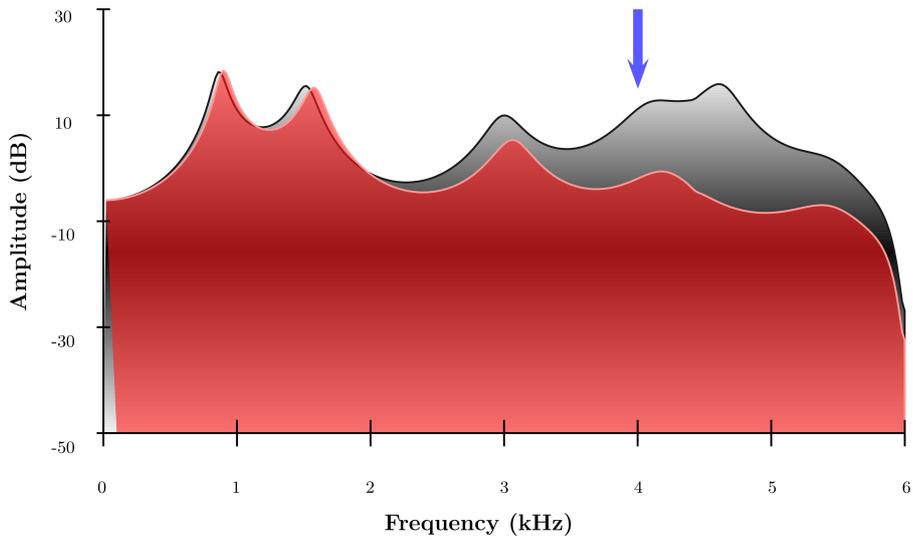


Figure 6.21: FEM simulated transfer function of VTM-MRI-Sophy-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

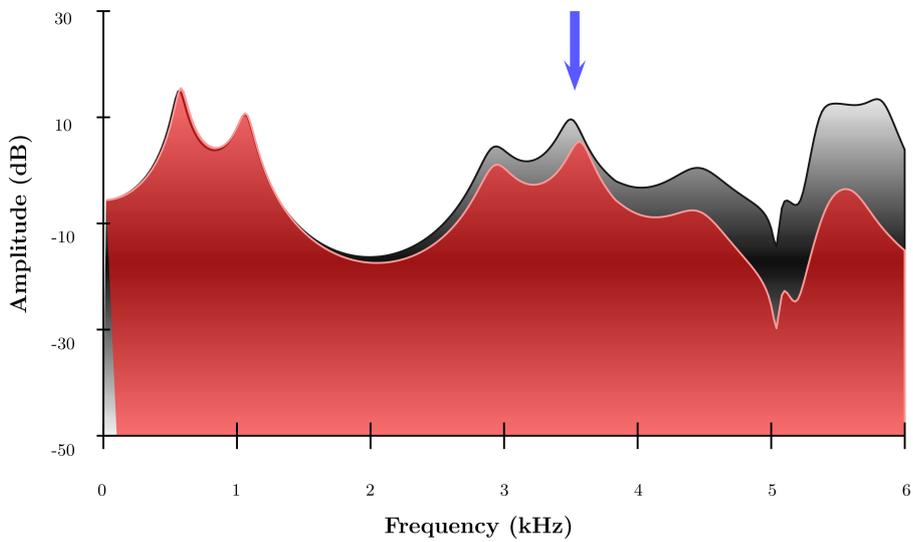


Figure 6.22: FEM simulated transfer function of VTM-MRI-Marisa-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

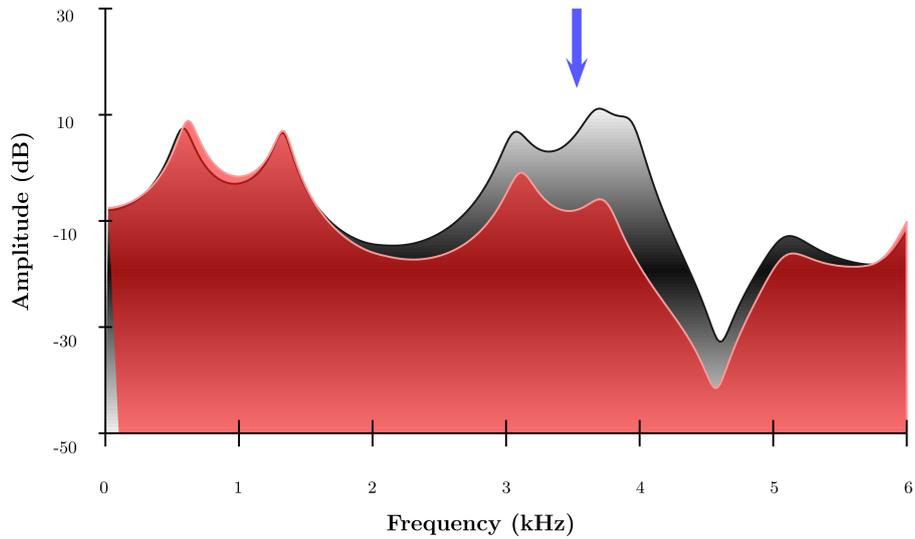


Figure 6.23: FEM simulated transfer function of VTM-MRI-Maristela/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

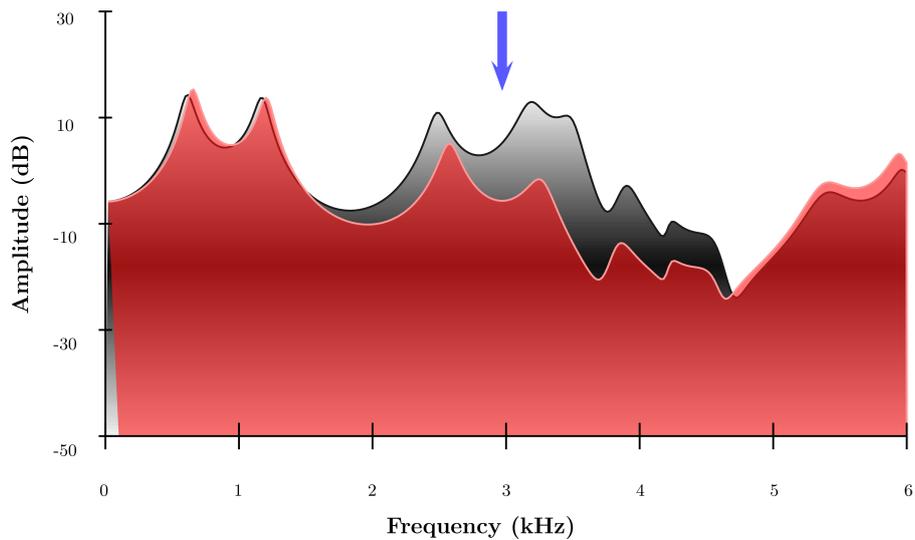


Figure 6.24: FEM simulated transfer function of VTM-MRI-Timothy/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

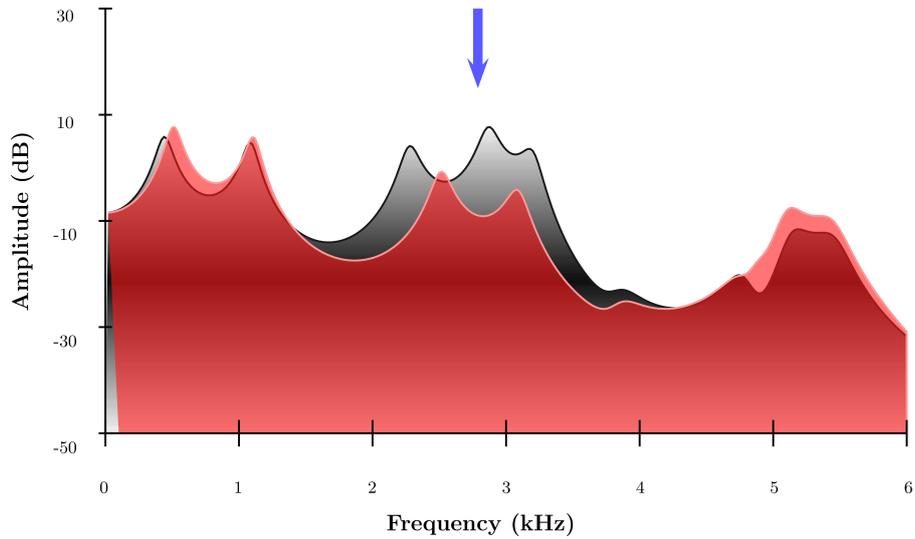


Figure 6.25: FEM simulated transfer function of VTM-MRI-Bartholomew-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

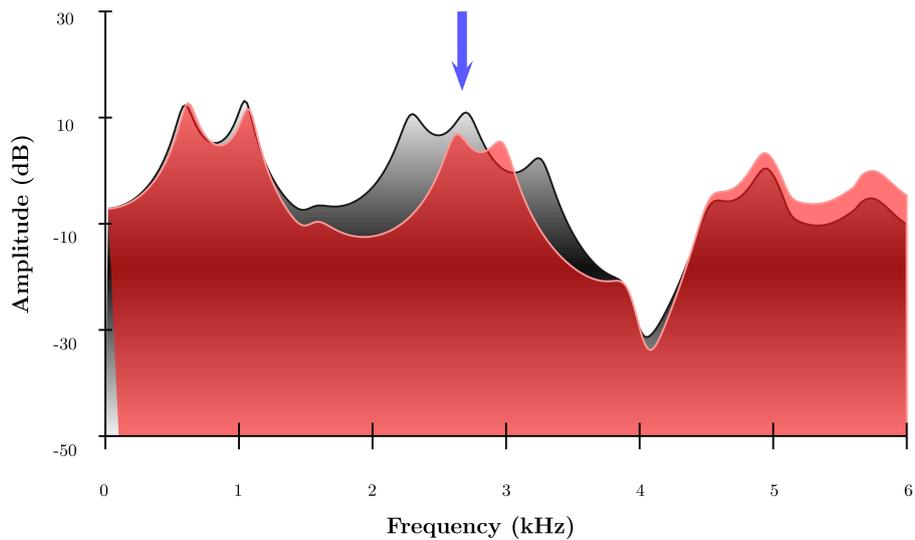


Figure 6.26: FEM simulated transfer function of VTM-MRI-Barnaby-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

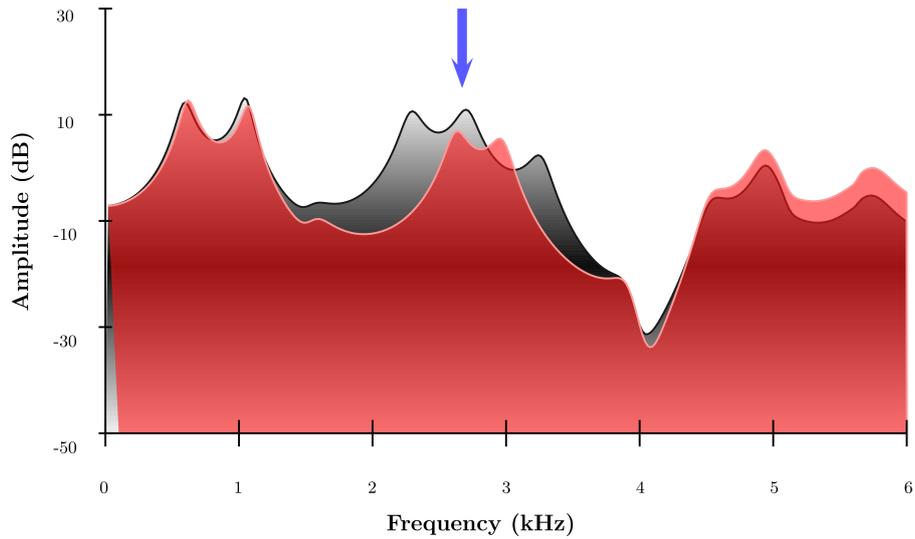


Figure 6.27: FEM simulated transfer function of VTM-MRI-Barnaby-/hard/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

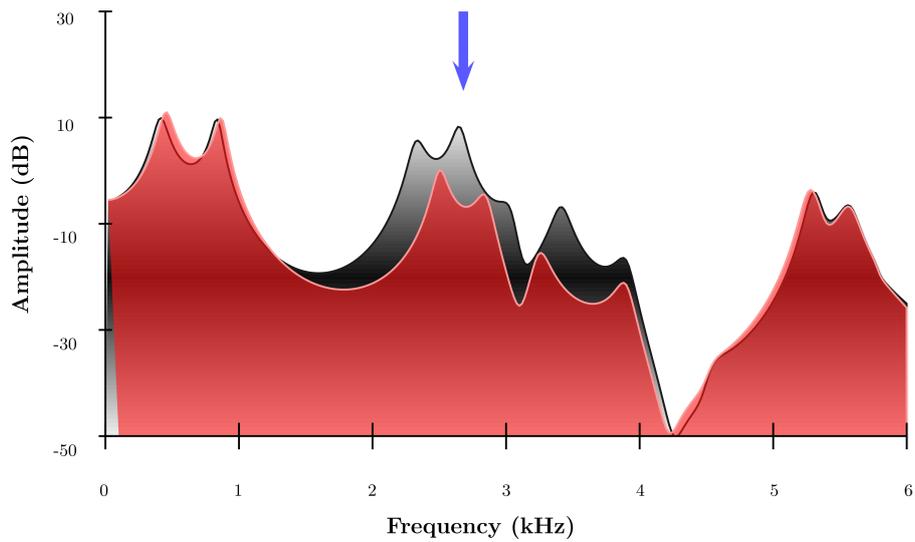


Figure 6.28: FEM simulated transfer function of VTM-MRI-Barnaby-/port/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

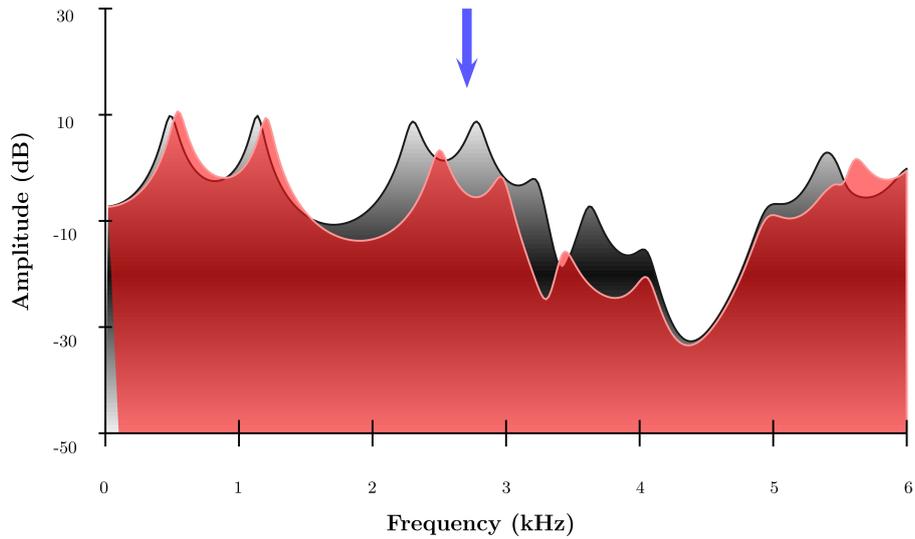


Figure 6.29: FEM simulated transfer function of VTM-MRI-Barnaby-/stern/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

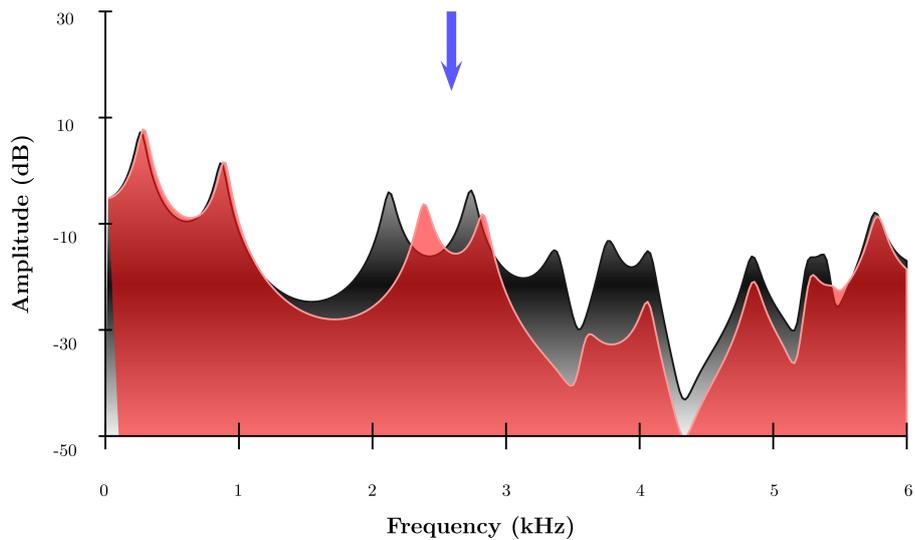


Figure 6.30: FEM simulated transfer function of VTM-MRI-Barnaby-/food/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

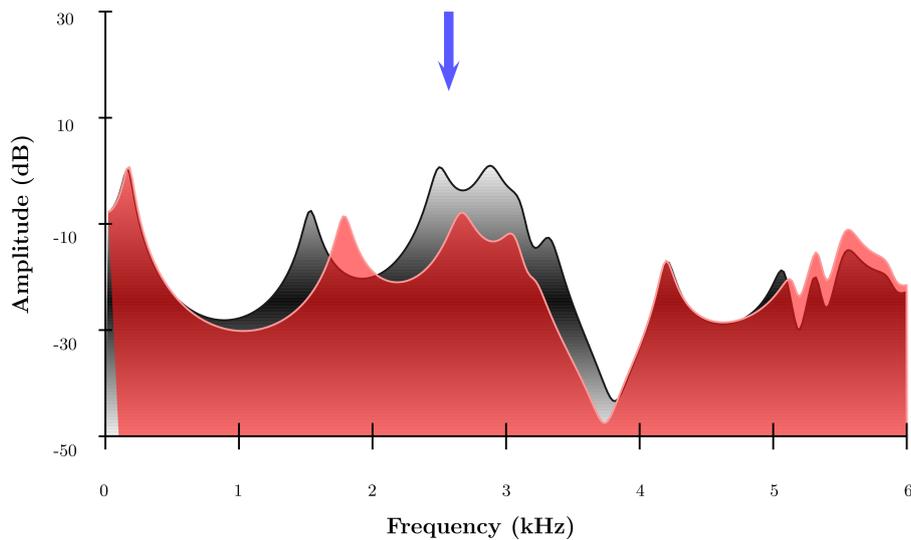


Figure 6.31: FEM simulated transfer function of VTM-MRI-Barnaby-/neap/, with (grayscale) and without (red) epilarynx. The blue arrow indicates the SFC centre predicted by the formula (6.12), derived from the dimensions of the hypopharynx.

From the dataset of six professional singers (see Figs 6.21, 6.22, 6.23, 6.24, 6.25, 6.27, 6.28, 6.29, 6.30, 6.31, the following was observed:

- Appending the epilaryngeal tube to the VTp creates an extra formant and locally shapes the neighbouring formant.
- The SFC frequency centre derived from the hypopharynx dimensions (blue arrow) provides a good prediction of the frequency around which the spectrum is locally shaped (*the focal point in the spectrum*, from Titze et al. [108]).
- Local resonances of the VTp smaller (resp. greater) than the predicted SFC frequency centre decrease (resp. increase) in the VT and an extra resonance arises in between.

This is in accordance with the results previously introduced when the epilaryngeal tube is appended to the oropharyngeal tube (see sections 6.1, 6.3.2).

The dataset of six professional singers provided in this thesis showed a correlation between SFC frequency centre, the epilarynx length and voice type: a longer epilarynx is synonymous with a lower SFC frequency centre and a lower voice type (see Table 6.2 and Fig 6.19). From the suggested formula (6.12) which predicts the SFC frequency centre from the hypopharynx dimensions, it is hypothesised that:

- **A greater epilaryngeal length leads to a lower SFC and hence a lower vocal fach**
- **To a lesser extent:**
 - A wider epilaryngeal cross-section at the pharyngeal opening increases the acoustical length of the epilaryngeal tube, lowering the SFC frequency centre, to the expense of the amplitude of the SFC (see section 6.3.2).
 - A wider pharyngeal opening increases the acoustical length of the epilaryngeal tube, lowering the SFC frequency centre.

6.4.3.2 Effect of lengthening the epilaryngeal tube

Figs 6.32, 6.33, 6.34, 6.35, 6.36 and 6.37 show the effect of lengthening the epilaryngeal tube on VTM-MRI.

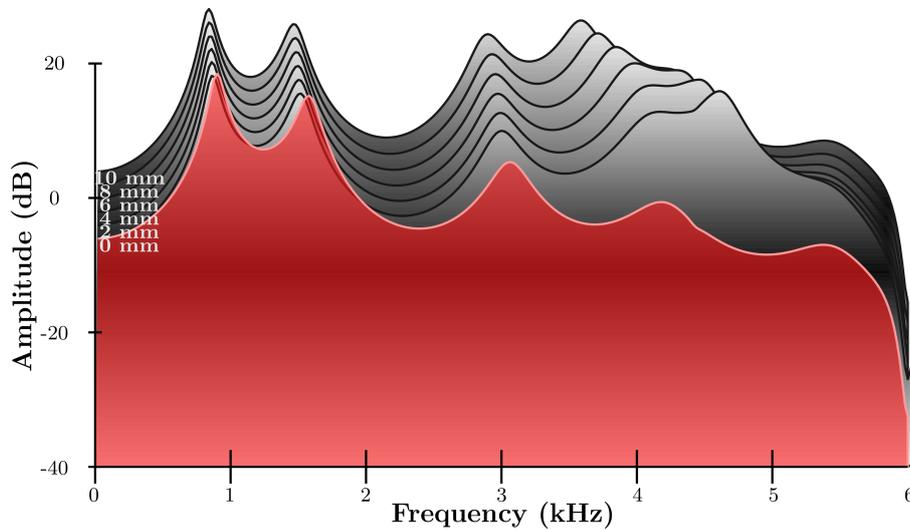


Figure 6.32: FEM simulated transfer functions of VTM-MRI-Sophy-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation.

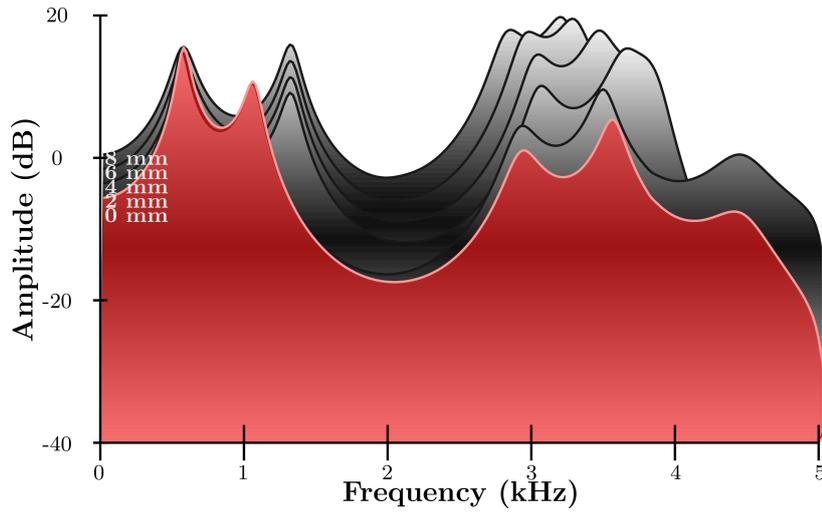


Figure 6.33: FEM simulated transfer functions of VTM-MRI-Marisa-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation.

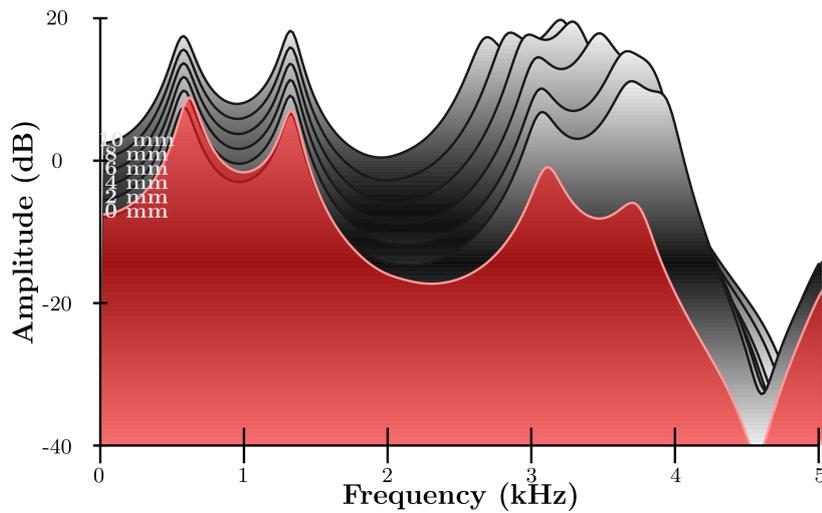


Figure 6.34: FEM simulated transfer functions of VTM-MRI-Maristela-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation.

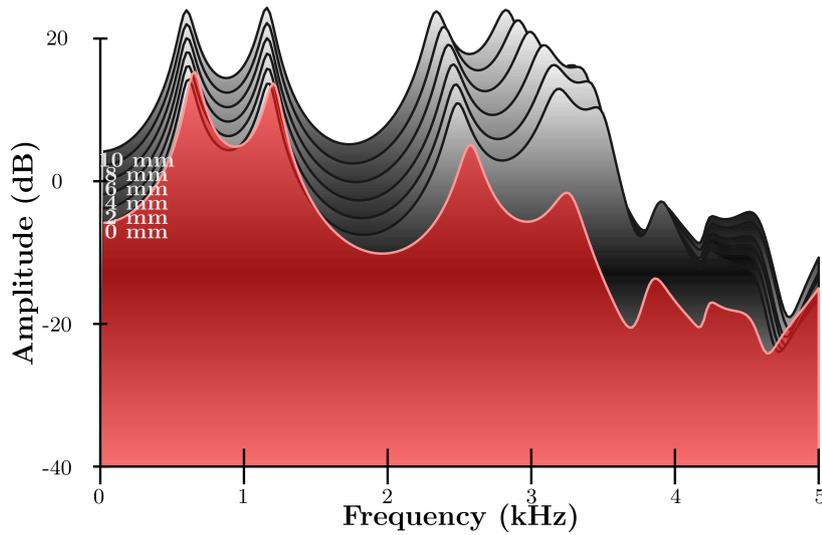


Figure 6.35: FEM simulated transfer functions of VTM-MRI-Timothy/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation.

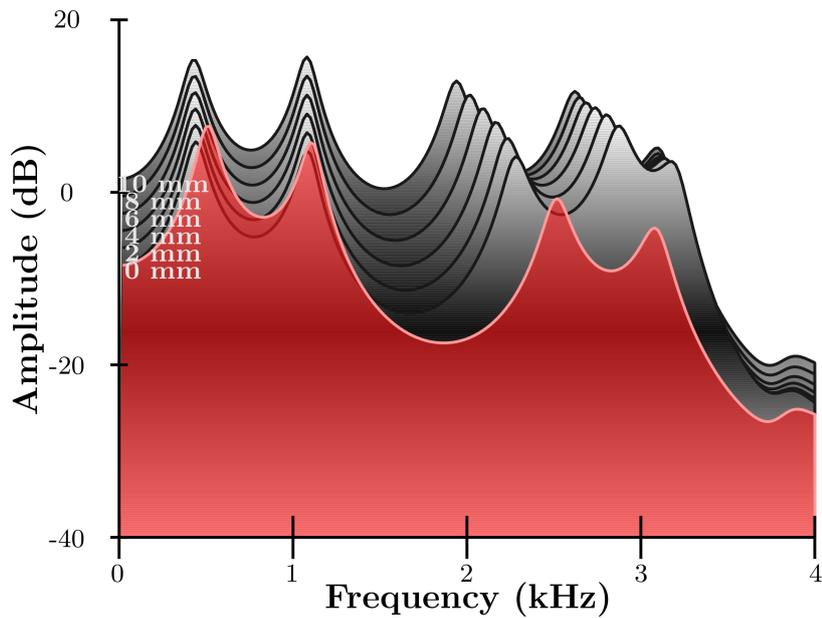


Figure 6.36: FEM simulated transfer functions of VTM-MRI-Bartholomew/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation.

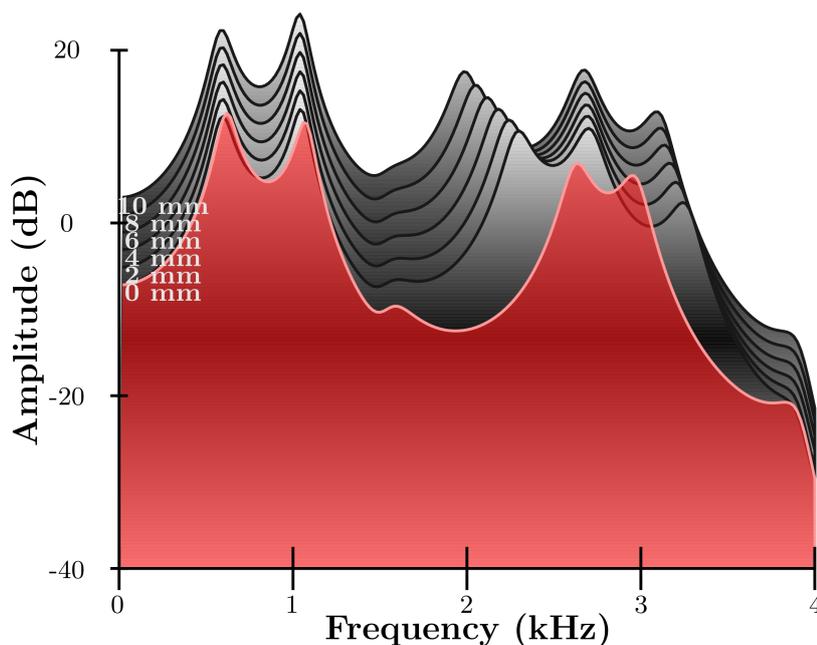


Figure 6.37: FEM simulated transfer functions of VTM-MRI-Barnaby-/hard/, with (greyscale) and without (red) the epilaryngeal tube when it is lengthened by 2 mm step downwards from 0 mm to 10 mm. Each step is plotted +2 dB from the previous one, for a better visualisation.

When the epilaryngeal tube is lengthened downwards by 2 mm steps, from 0 mm to 10 mm, R1 and R2 remain essentially the same, i. e. the vowel quality remains constant, whereas the higher resonance frequencies, such as R3, R4 and R5 are globally shifting towards lower frequencies (see Table 6.3), resulting in a lower SFC, which in this dataset corresponds to a lower vocal fach. This is in accordance with the results given in section 6.3.

The values given in Table 6.3 represent the shift (in %) of the resonant frequencies when the epilarynx is lengthened downwards from 0 mm to 10 mm (these are the values of the peaks from Figs 6.32, 6.33, 6.34, 6.35, 6.36 and 6.37). The last column gives the shift of the SFC frequency centre computed with the new metric (see section 6.5).

Resonance frequencies shift (in %)						
	R ₁	R ₂	R ₃	R ₄	R ₅	SFC
Sophy	-3.23	-3.04	-3.46	-13.77	-9.5	-13.52
Marisa	-1.61	-0.28	-0.44	-3.43	-1.23	(1.74)
Maristela	-0.68	0	-12.45	-14.30	-3.09	-12.88
Timothy	-2.12	-1.03	-5.99	-11.5	-5.27	-10.02
Bartholomew	-2.71	-0.37	-14.99	-8.77	-2.55	-13.57
Barnaby	-1.17	-0.57	-13.61	-0.96	-4.66	-8.98
(μ)	(-1.92)	(-0.88)	(-8.49)	(-8.79)	(-4.38)	(-9.54)
μ	-1.98	-1	-10.1	-9.86	-5.02	-11.79
(σ)	(0.95)	(1.11)	(6)	(5.52)	(2.9)	(5.85)
σ	1.05	1.19	5.06	5.43	2.74	2.14
	R ₁ &R ₂		R ₃ &R ₄ &R ₅			
μ	-1.49		-8.33			

Table 6.3: Resonance frequencies shift (in%) when the epilarynx is lengthened downwards by 10 mm. Note that Marisa's MRI data showed some artifacts and were therefore more difficult to segment accurately: this is why the data obtained by lengthening the segmented epilarynx are not reliable and therefore put in between brackets.

Note that Marisa's MRI data showed some artifacts and were therefore more difficult to segment accurately: this is why the data obtained by lengthening the segmented epilarynx are not reliable and therefore put in between brackets.

The SFC frequency centre shows a clear trend of decreasing its frequency when the epilarynx is lengthened downwards, with an average of 11.79 % decrease (without Marisa). This trend is confirmed graphically with the Figs 6.38, 6.39, 6.40, 6.41, 6.42 and 6.43. These Figs represent the view from the top of Figs 6.32, 6.33, 6.34, 6.35, 6.36 and 6.37) respectively. They show the 5 first resonant frequencies varying with the lengthening of the epilarynx from 0 mm to 10 mm (in grey) and the SFC frequency centre computed with the new metric from section 6.5 (in purple). Observe that the formants and SFC describe different patterns of shifting depending on the singers. For instance, for Sophy (Fig 6.38), R4 and R5 get closer to each other for an epilaryngeal lengthening from 0 mm to 4 mm where they cluster, before diverging again as the epilarynx is lengthened further. In the case of Maristela (Fig 6.40), R4 and R5 cluster at a lengthening of 0 mm, coincide at 2 mm and then diverge with R3 getting closer and clustering with R2. The SFC (in purple) follows the clustering of the resonances accordingly. For Timothy (Fig 6.41), the SFC follows essentially the path of R4 (which is the dominant shift in the higher resonances: -11.5 % versus -5.99 % for R3 and -5.27 % for R5). R4 and R5 converge until a lengthening of 4 mm before diverging. In the case of Bartholomew (Fig 6.42), the SFC follows a path parallel to those of R3 and R4, which have a respective shift of -14.99 % and -8.77 % against only -2.55 % for R5. The case of Barnaby (Fig 6.43) shows a lesser SFC shift than the other singers (-8.98 % versus values between -10 % and -13.57 %) with a clear shift of R3 (-13.61 %), to a lesser extent R5 (-4.66 %) and a stable R4 (-0.96 %) which results in a diverging R3-R4 and converging R4-R5.

In Table 6.3, it can be seen that R₁ and R₂ remain essentially constant when the epilarynx is lengthened with a average of -1.98 % and -1 % respectively, whereas the higher resonances R₃, R₄ and R₅ follow more significant changes, -10.1 %, -9.86 % and -5.02 % respectively. The average shifting for R₁ and R₂ is -1.49 % to be compared with -8.33 % for the higher resonances. This is in accordance with the results given in section 6.3: a lengthening of the epilarynx does not affect greatly F₁ and F₂, maintaining the frequency content which leads to the vowel identification, whereas the higher formants F₃, F₄ and F₅ vary greatly and change therefore the vowel quality as well as the SFC frequency centre, and therefore the voice classification (see Fig 6.19).

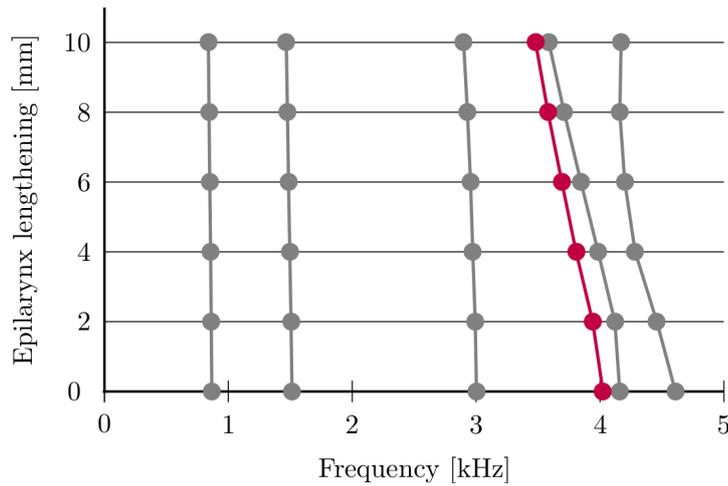


Figure 6.38: Five first resonance frequency values for VTM-MRI-Sophy-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.32 from the top.

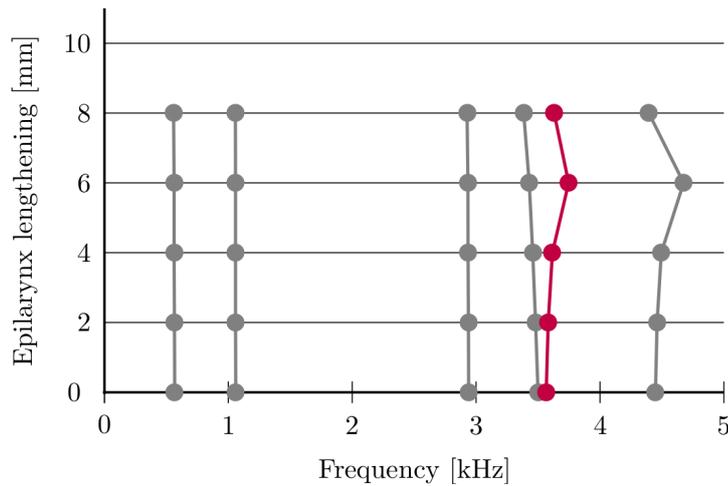


Figure 6.39: Five first resonance frequency values for VTM-MRI-Sophy-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.33 from the top.

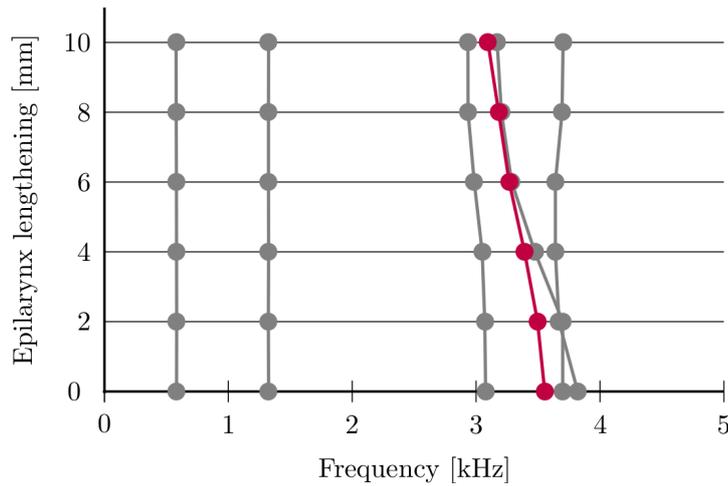


Figure 6.40: Five first resonance frequency values for VTM-MRI-Maristela/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.34 from the top.

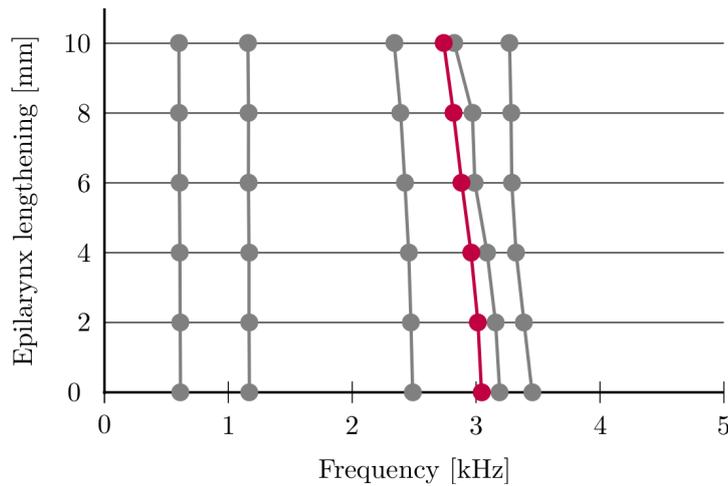


Figure 6.41: Five first resonance frequency values for VTM-MRI-Timothy/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.35 from the top.

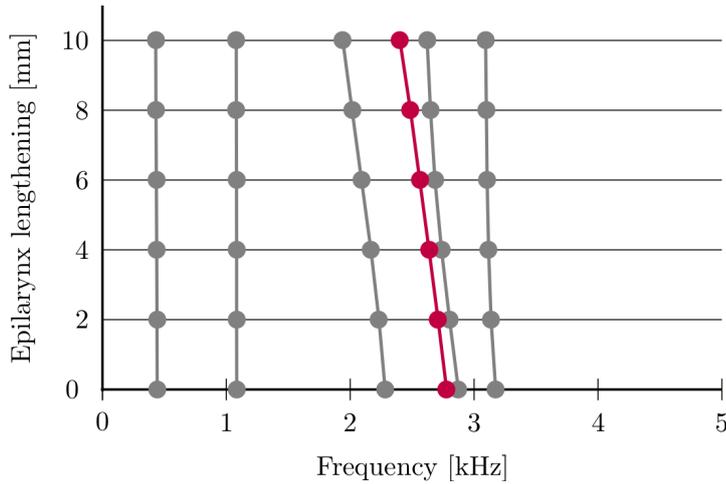


Figure 6.42: Five first resonance frequency values for VTM-MRI-Bartholomew-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.36 from the top.

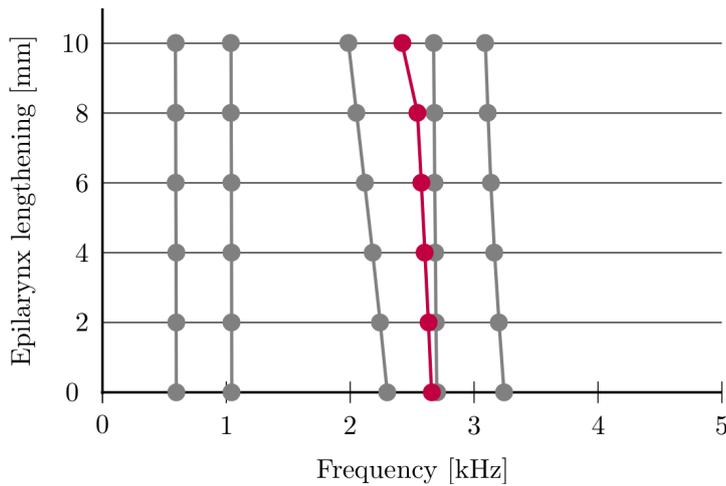


Figure 6.43: Five first resonance frequency values for VTM-MRI-Barnaby-/hard/, when the epilarynx is lengthened by 2 mm step downwards from 0 mm to 10 mm. In purple, the SFC frequency centre computed from the new metric, in section 6.5. This is the view of 6.37 from the top.

6.4.4 Effect of the piriform fossae

Appending the piriform fossae to the VTp + epilaryngeal tube, (two “peer-shaped” pockets located posteriorly at the bottom of the pharynx) adds a trough around 4-5 (6) kHz in the output spectrum, probably enhancing the perception of the SFC: a broad peak, followed by a trough [48]. Similar findings are visually illustrated on Figs 6.44, 6.45 experimentally and on Figs 6.46, 6.47, 6.48, 6.49, 6.50, 6.51 and Figs 6.51, 6.52, 6.53, 6.54, 6.55 numerically for the 6 singers and Barnaby on 5 vowels respectively.

On the Figs 6.44 and 6.45, the greyscale plots represent the transfer function derived from experimental measurement on VTM-MRI-Barnaby-/food/ and VTM-MRI-Barnaby-/neap/ respectively. For the experimental measurement, plasticine was used to fill the piriform fossae. It is difficult to manually smooth the plasticine in the piriform fossae and fill the entire gap, which is probably why Fig 6.45 exhibits a behaviour less clear than Fig 6.44 for instance. By filling the piriform fossae, their spectral effect is removed from the overall transfer function: the red (blue) plot represents the VTM-MRI without the left (right) piriform fossa. Note that each piriform fossa has its own shape and dimensions, and it is very unlikely that the piriform fossae are completely symmetrical. Therefore, each piriform fossa acts as a side branch which brings its own antiresonance frequency, as can be seen on Figs 6.44 and 6.45 where both spectral zeros (troughs in the red and the blue plots) are distinct from each other. The blue and red spectral zeros in VTM-MRI-Barnaby-/food/ are closer than those from VTM-MRI-Barnaby-/neap/: this suggests that the piriform fossae of VTM-MRI-Barnaby-/food/ are more similar in shape and dimensions than those of VTM-MRI-Barnaby-/neap/. It can be seen that the main frequency region affected by the piriform fossae is between 4 and 5 kHz. The formants below and above this region are repelled: the formants whose frequencies are lower/greater than the

resonance frequency of the fossae are decreased/increased respectively when the fossae are appended to the tract. This agrees with the results found in [48, 18, 21, 108].

Figs 6.46, 6.47, 6.48, 6.49, 6.50, 6.51 and Figs 6.51, 6.52, 6.53, 6.54, 6.55 show the simulated transfer functions (FEM) for the 6 singers and Barnaby on 5 vowels respectively. The green arrow represents the resonance frequency of the piriform fossae derived from their length (see Table 6.4). Titze et al. suggested to use the quarter-wave resonator formula (eq (13) from [108])

$$F_{sn} = (2n - 1) \frac{c}{4L_s} \quad (6.13)$$

where F_{sn} is the n^{th} resonance of the piriform sinuses, c the speed of sound and L_s the length of the sinuses. The predicted spectral zeros are in good accordance with the numerical simulations: the longer the sinus, the lower the resonance frequency. Knowing more accurately the acoustical length of the fossae (accounting for the end correction effect) would give a more accurate prediction of the resonance frequency.

The mean antiresonance frequency across singers is 5057 Hz with a standard deviation of 1037 Hz whereas the mean value across the vowels of Barnaby is 4182 Hz with a standard deviation of 179 Hz. The average length of the piriform sinuses across singers is 17.10 mm with a standard deviation of 2.83 mm whereas this average is 20.01 mm with a standard deviation of 0.86 mm across the vowels of Barnabas. The data show naturally more consistency and less variation intraindividually than interindividually. The variability differs amongst genders, with a standard deviation for the length piriform sinuses for females of 2.98 mm versus 1.50 mm for the males, which gives a ratio 2:1. The resonance frequency of the fossae gives 1135 Hz for the females versus 273 Hz for the males. Figs 6.46, 6.47, 6.48, 6.49, 6.50, 6.51 and Figs 6.51, 6.52, 6.53, 6.54, 6.55 visually confirm the experimental results: the piriform fossae act as formants repellents, the formants

with a lower/greater frequency than the resonance frequency (green arrow) see their frequency decreased/increased.

Vocal Tract and piriform fossae dimensions					
	L_s	f	VTV	PV	%
Sophie - hard	12	6965	82.54	1.7	2.06
Maristela - hard	17.63	4741	44.58	3.47	7.78
Marisa - hard	16.5	5065	111.82	1.8	1.61
Timothy - hard	17.63	4741	67.26	2.51	3.73
Bartholomew - hard	18.31	4565	37.32	1.7	4.56
Barnaby - hard	20.50	4077	80.49	1.69	2.1
μ	17.10	5057	70.67	2.15	3.63
σ	2.83	1037	27.33	0.72	2.32
μ (females)	15.38	5723	79.65	2.32	3.82
σ (females)	2.98	1135	33.71	0.99	3.44
μ (males)	18.81	4392	61.69	1.97	3.46
σ (males)	1.50	273	22.11	0.47	1.25
Barnaby - hard	20.50	4077	80.49	1.69	2.1
Barnaby - port	19.97	4185	65.61	3.64	5.55
Barnaby - stern	19.00	4399	65.29	3.43	5.25
Barnaby - food	19.42	4304	57.67	4.79	8.31
Barnaby - neap	21.18	3946	65.48	3.58	5.47
μ (vowels)	20.01	4182	63.48	3.42	5.33
σ (vowels)	0.86	179	9.41	1.11	2.20

Table 6.4: L_s is the averaged length of the two piriform sinuses [mm], f, the antiresonance frequency generated by the sinuses [Hz], VTV is the volume of the Vocal Tract [cm³], PV is the volume of the piriform fossae [cm³], % is the ratio PV/VTV expressed in percentage. The corresponding statistical average (μ) and standard deviation (σ) are also given.

It is interesting to note that the ratio of the volume of the piriform fossae and the Vocal Tract (penultimate column in Table 6.4) is related to the amplitude of their effect on the spectrum: the bigger the fraction, the bigger the impact on the transfer function. See for instance Maristela, whose piriform fossae constitute 8% of the Vocal Tract volume: her piriform fossae have a relatively larger spectral impact than those of the other singers.

From Figs 6.46, 6.47, 6.48, 6.49, 6.50 and 6.51, it can be seen that the female voice tends to show a spectral trough due to the piriform fossae at a higher frequency range (around 4-5 to 6 kHz) than males (around 3.5 to 5 kHz), which is consistent with the fact that the spectral role of the piriform fossae is to emphasise the SFC.

Moreover, the physiological role of the piriform fossae is to serve as side branches to "capture" foreign bodies, instead of swallowing them, but also a part of the food (at least temporarily) and the mucus, for instance when one has a cold [3]. It is suggested, therefore, that singers with large piriform fossae may be more affected than others in the production of a "resonant voice" when they have a cold or when they have just eaten certain foods which would obstruct the fossae.

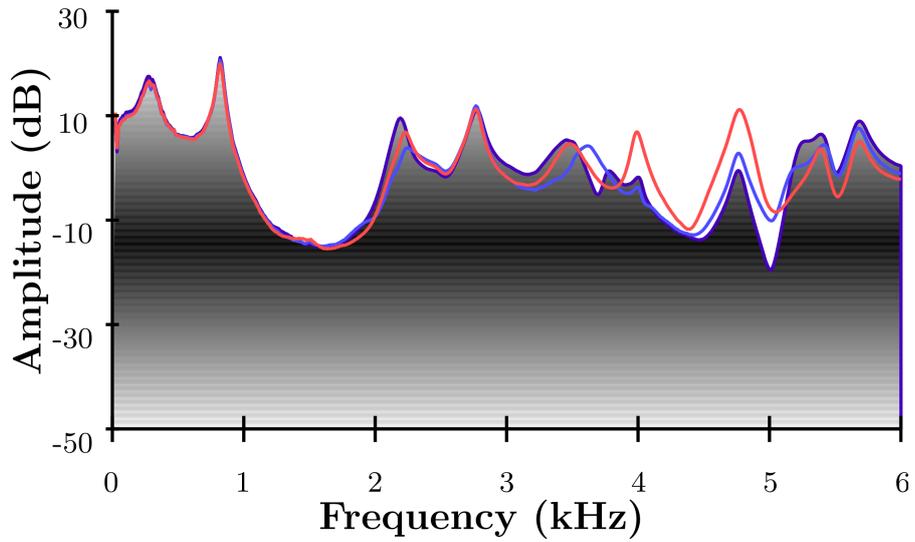
6.4.4.1 *Experimental results*

Figure 6.44: Experimental results for VTM-MRI-Barnaby-/food/ with (grayscale) and without left (red) or right (blue) piriform fossa.

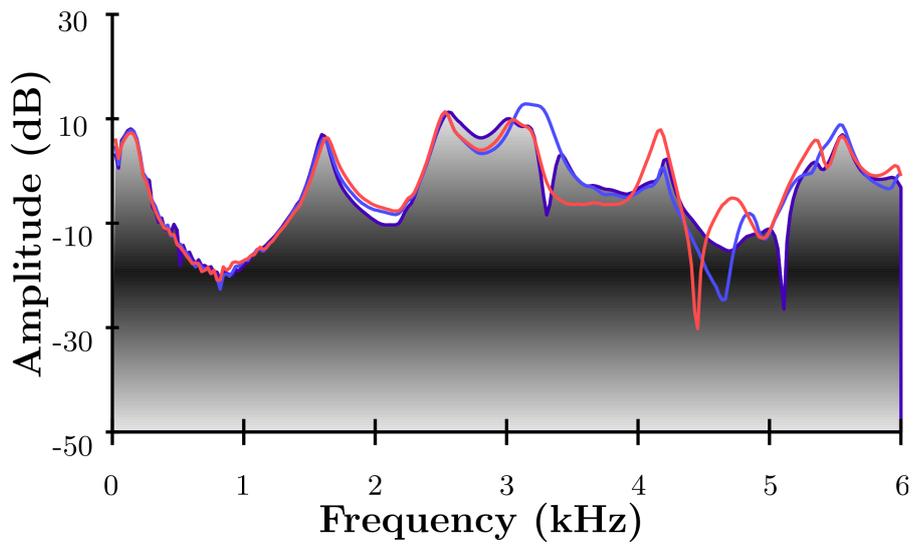


Figure 6.45: Experimental results for VTM-MRI-Barnaby-/neap/ with (grayscale) and without left (red) or right (blue) piriform fossa.

6.4.4.2 Numerical simulations

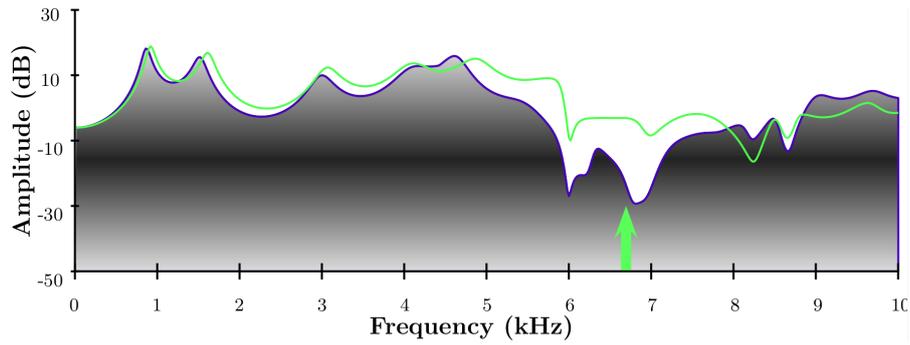


Figure 6.46: Numerical results for VTM-MRI-Sophy-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

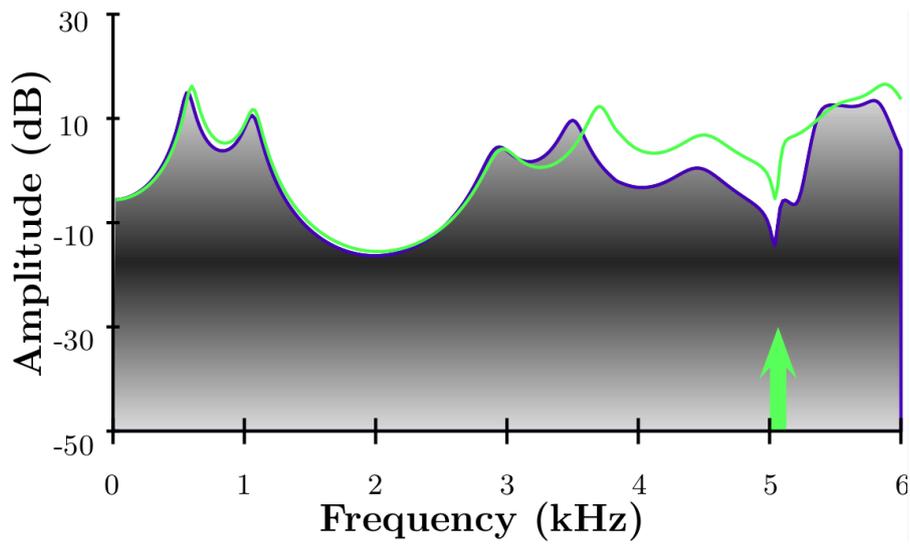


Figure 6.47: Numerical results for VTM-MRI-Marisa-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

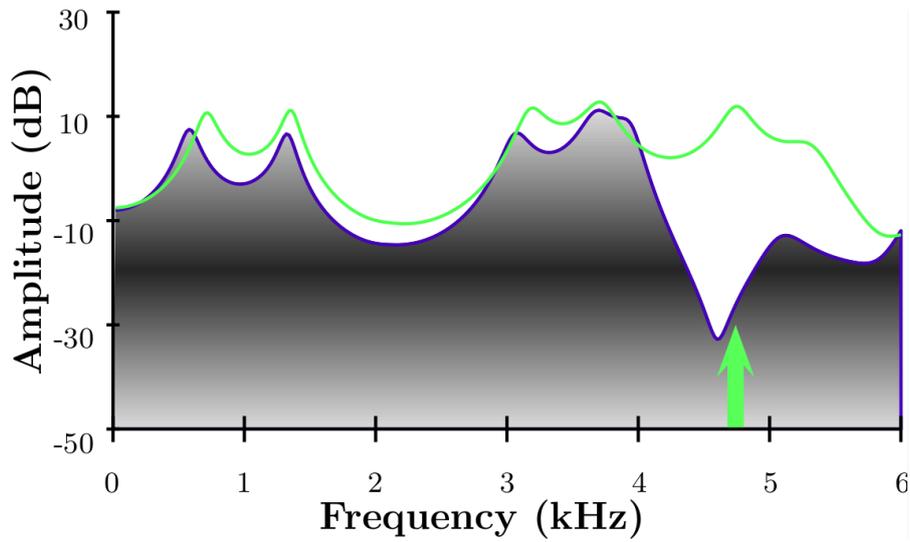


Figure 6.48: Numerical results for VTM-MRI-Maristela-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

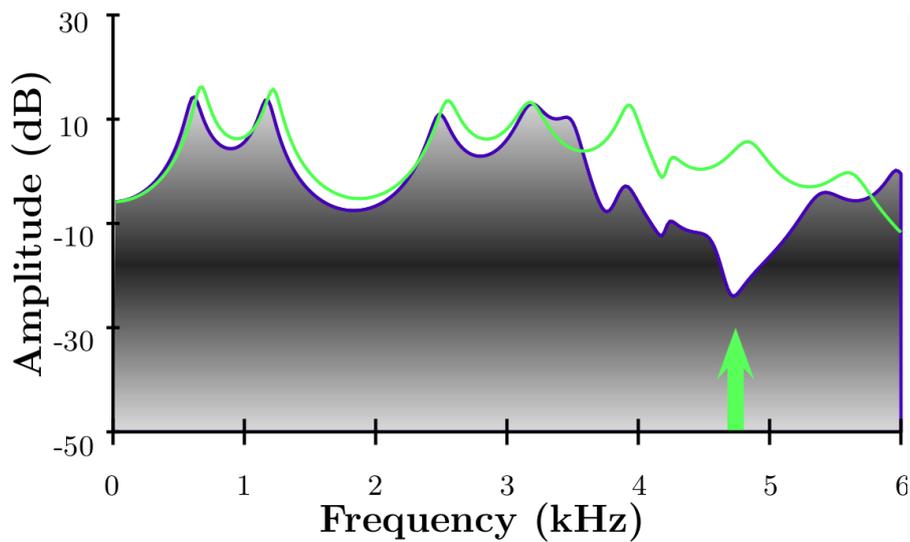


Figure 6.49: Numerical results for VTM-MRI-Timothy-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

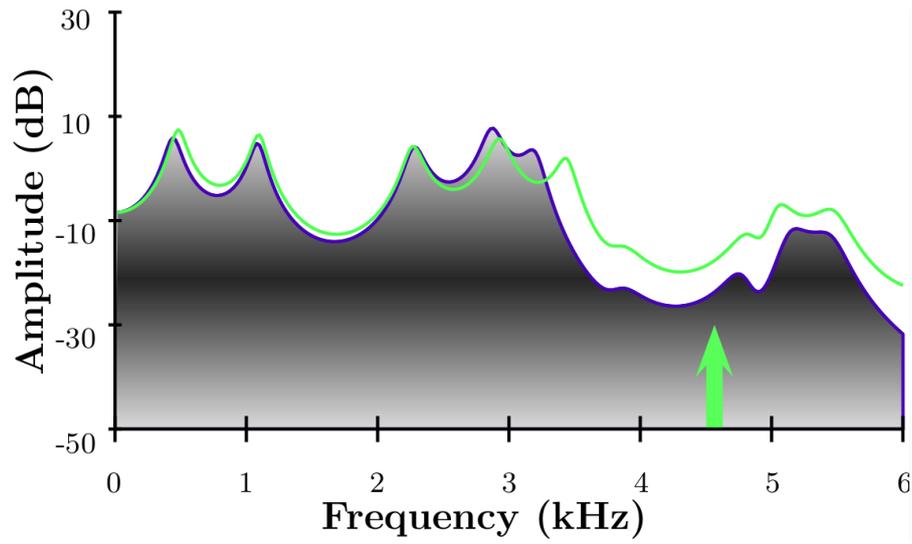


Figure 6.50: Numerical results for VTM-MRI-Bartholomew-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

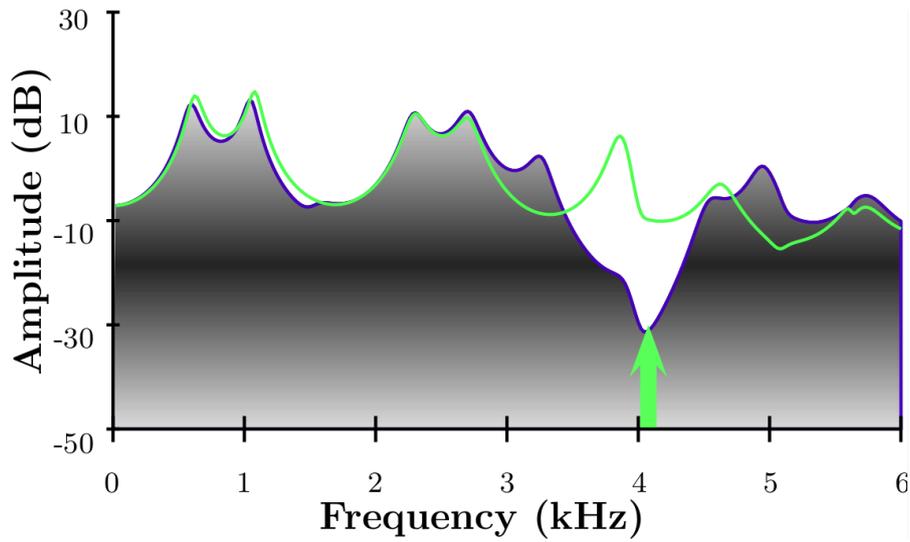


Figure 6.51: Numerical results for VTM-MRI-Barnaby-/hard/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

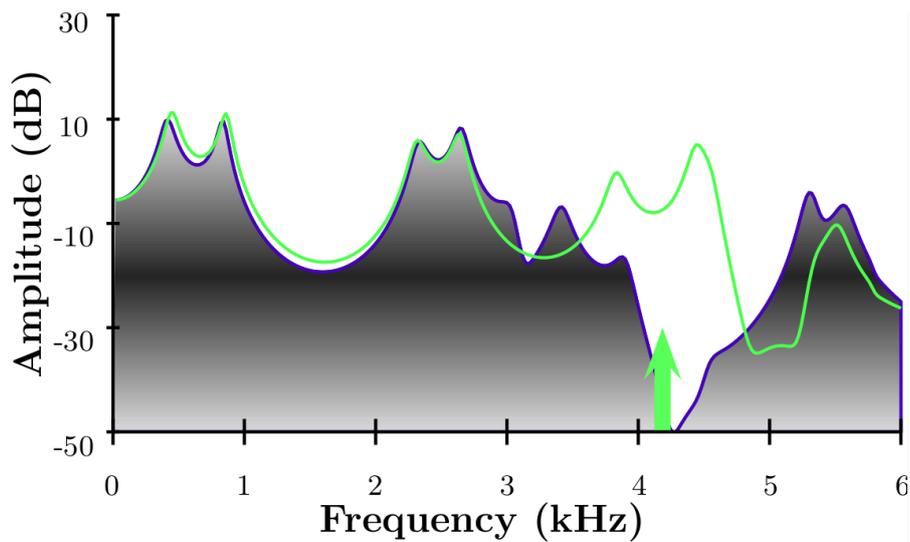


Figure 6.52: Numerical results for VTM-MRI-Barnaby-/port/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

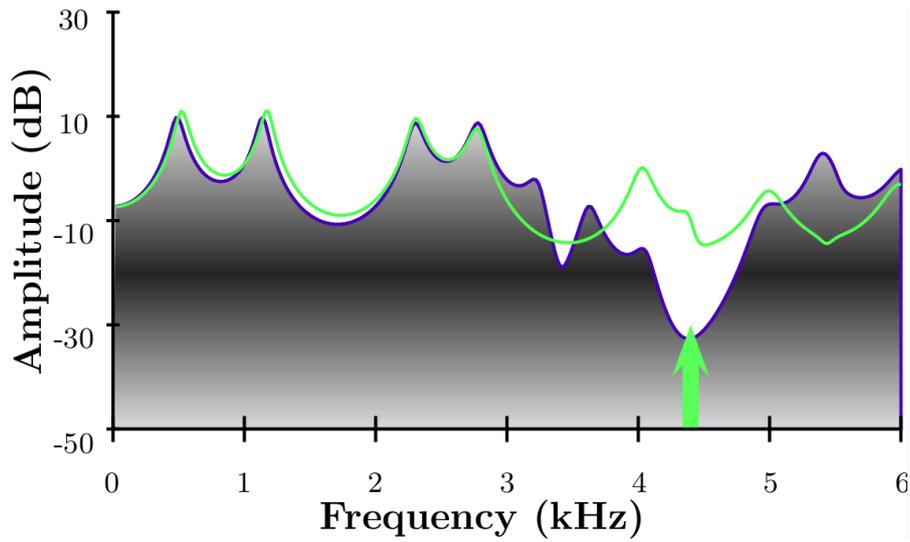


Figure 6.53: Numerical results for VTM-MRI-Barnaby-/stern/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

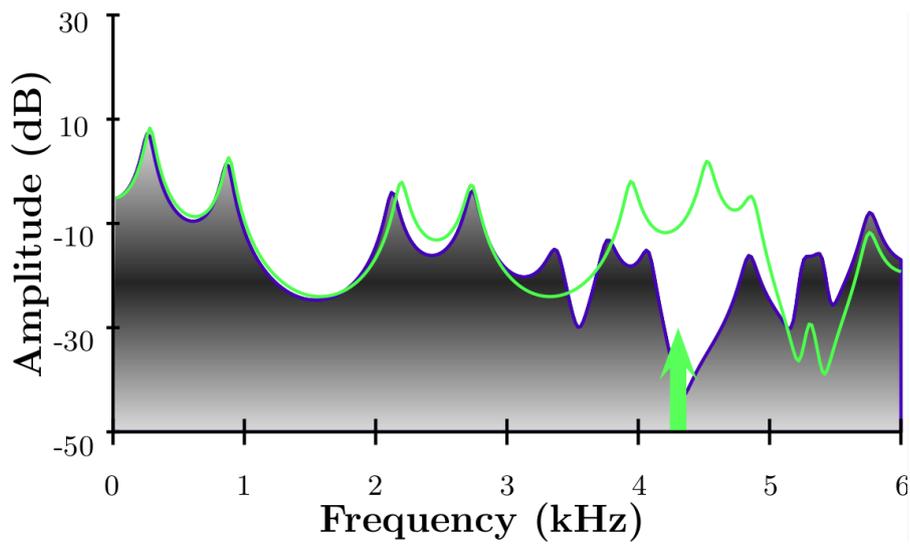


Figure 6.54: Numerical results for VTM-MRI-Barnaby-/food/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

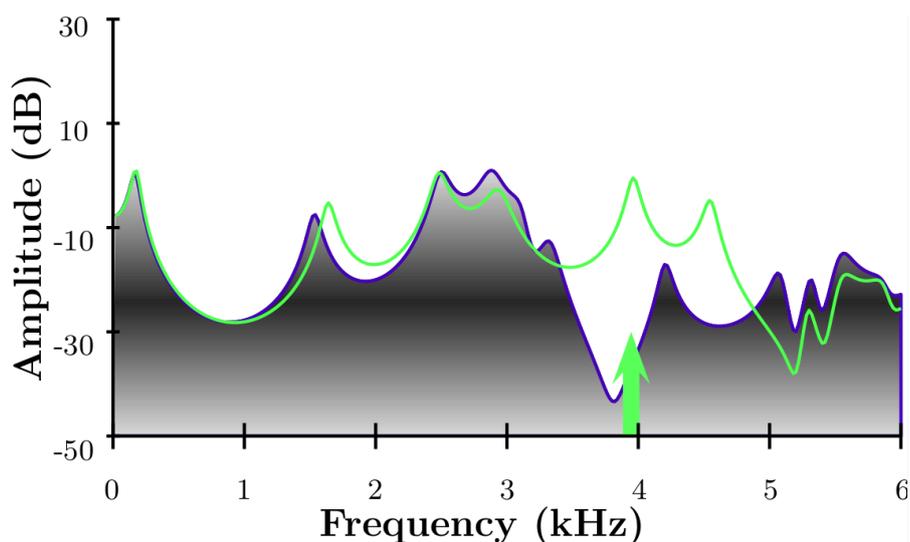


Figure 6.55: Numerical results for VTM-MRI-Barnaby-/neap/ with (grayscale) and without (green) piriform fossae. The green arrow represents the antiresonance frequency of the piriform fossae.

6.4.5 Effect of the vallecula

The effect of the vallecula seems to enhance the SFC by accentuating the dip created by the piriform fossae (Figs 6.56, 6.57 from Barnaby phonating on /neap/ and /food/ respectively). The spectral effect of the vallecula could only be assessed on the 3D-printed VTM-MRI (through the placement of plasticine inside the corresponding cavities) because the twisted structure of the vallecula prevented BLENDER from any mesh manipulation without severe alteration. Like the piriform fossae, they act as side branches to the main tract, damping a frequency range around 4-5 kHz.

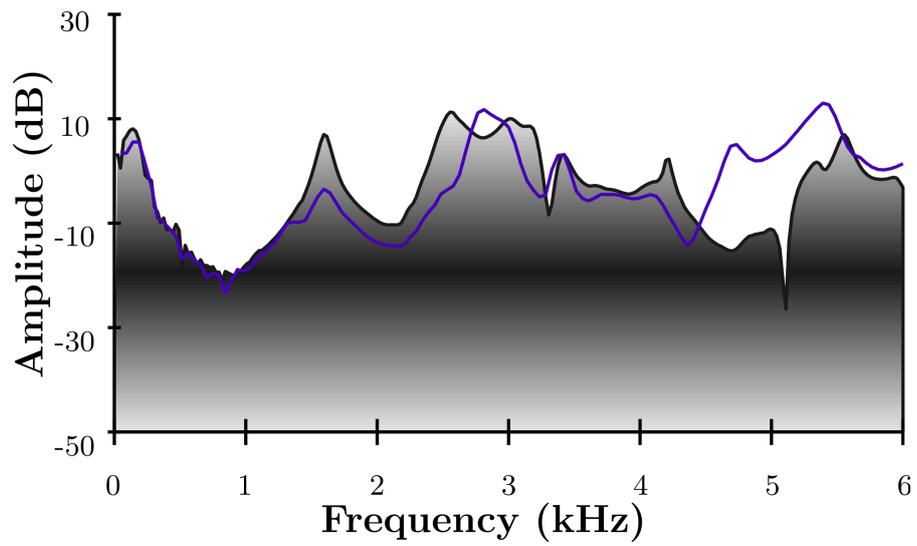


Figure 6.56: Experimental results of MRI-based 3D-printed VT of Barnaby, phonating on /neap/ with (greyscale) and without (blue) val- lecula.

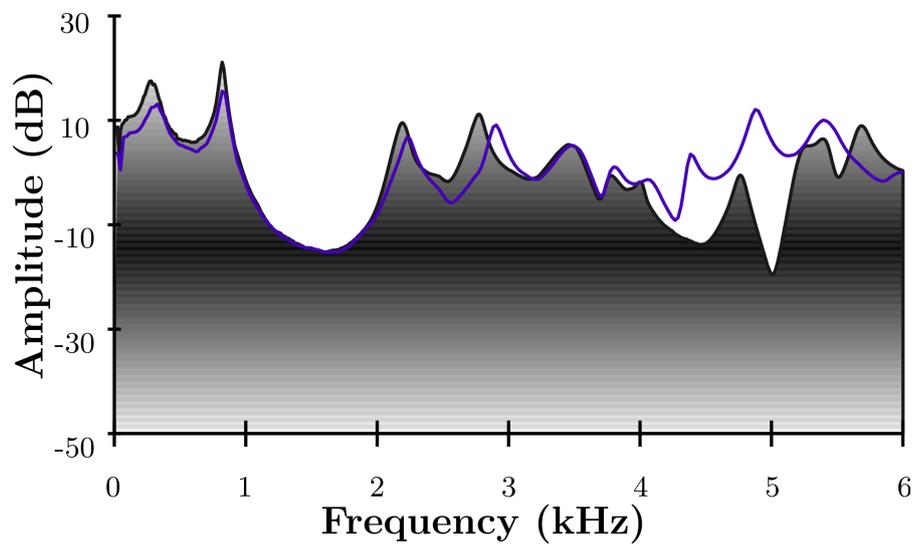


Figure 6.57: Experimental results of MRI-based 3D-printed VT of Barnaby, phonating on /neap/ with (greyscale) and without (blue) val- lecula.

6.4.6 Numerical versus experimental

Figs 6.58, 6.59, 6.60, show the comparison between numerical simulations (in blue) and experimental results (in red) for Barnaby phonating on /port/, /food/ and /stern/ respectively. The numerical simulations match rather well the experimental results, although there are differences, mostly due to the fact that the simulation propagates a lossless wave equation whereas the actual Vocal Tract implies fluid dynamics with turbulence, vorticity, viscous layers, heat losses, etc. Moreover, the absorption coefficient of the simulation is not frequency dependent.

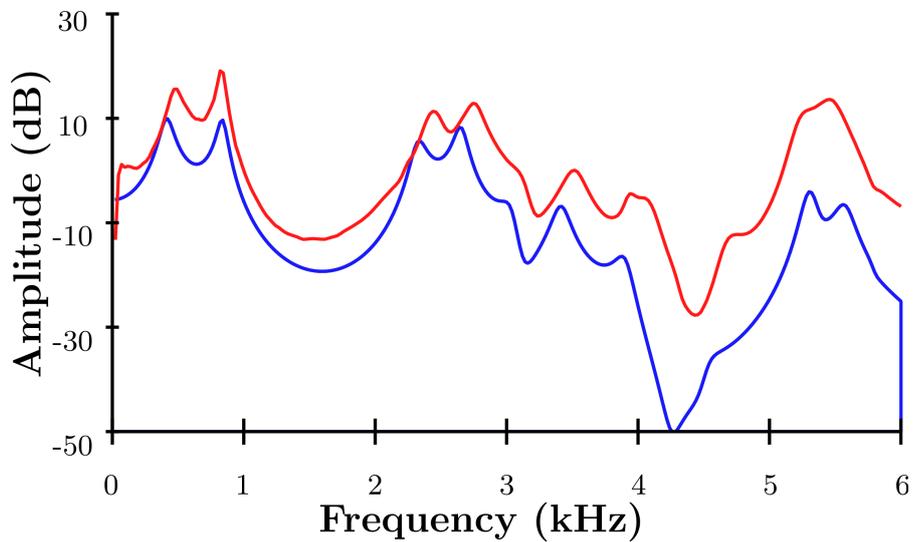


Figure 6.58: Numerical (blue) simulation versus experimental (red) results of Barnaby singing on /port/.

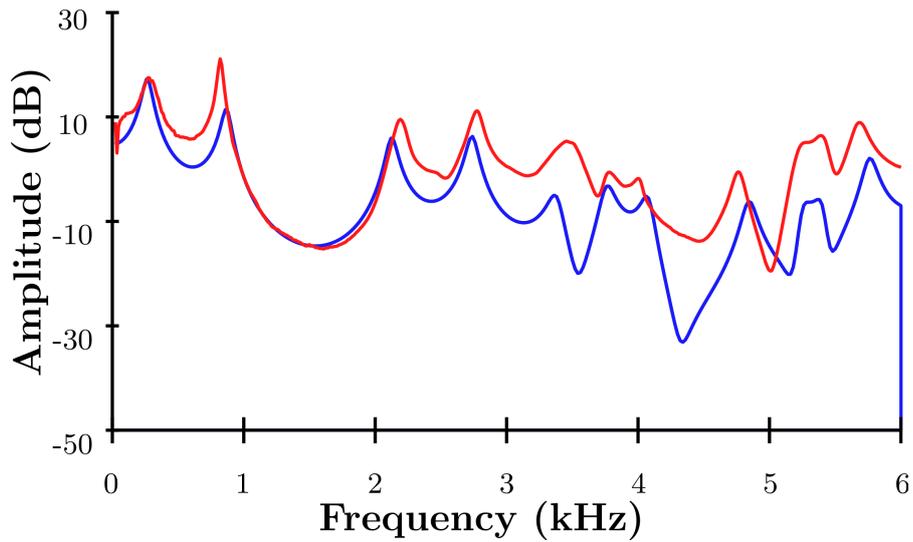


Figure 6.59: Numerical (blue) simulation versus experimental (red) results of Barnaby singing on /food/.

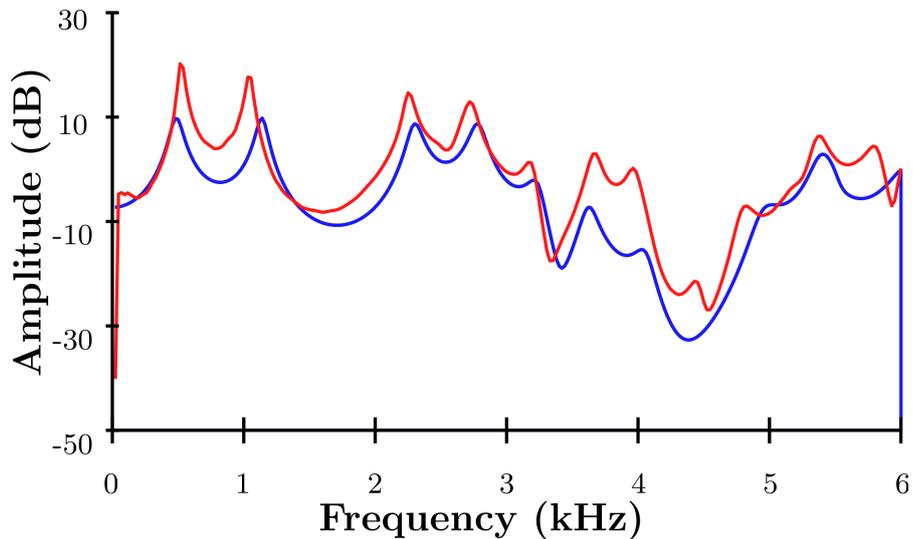


Figure 6.60: Numerical (blue) simulation versus experimental (red) results of Barnaby singing on /stern/.

Table 6.5 shows the comparison between the simulated and measured formant frequencies, and their relative difference for Barnaby. It is to be noted that, apart from the 3D-printed VT of Barnaby phonating on /hard/, the numerical results are in a good agreement with the experiment. VTM-MRI-Barnaby-/hard/ experienced a problem in the 3D-printing process (see the only peak for F₃-F₄-F₅); it needed to go through a completely new segmentation process for simula-

tion purposes. The numerical and experimental Vocal Tract shapes of Barnaby come from two different segmentations and lead naturally to discrepancies in the resonant frequencies.

Formant frequencies (in Hz): simulation versus experimental

	F1 (num)	F1 (exp)	%	F2 (num)	F2 (exp)	%	F3 (num)	F3 (exp)	%	F4 (num)	F4 (exp)	%	F5 (num)	F5 (exp)	
hard	596	616	3.25	1,040	1,068	2.62	2,299	2,743	16.19	2,700	2,743	1.57	3,243	2,743	-18.23
port	417	482	13.49	834	833	-0.12	2,334	2,448	4.66	2,647	2,753	3.85	3,414	3,515	2.87
stern	483	523	7.65	1,134	1,043	-8.72	2,302	2,256	-2.04	2,778	2,719	-2.17	3,625	3,173	-14.25
food	264	283	6.71	867	821	-5.60	2,122	2,193	3.24	2,736	2,774	1.37	3,361	3,464	2.97
neap	165	142	-16.20	1,531	1,604	4.55	2,500	2,560	2.34	2,879	3,019	4.64	3,313	3,410	2.84

Table 6.5: Comparison between the experimental and numerical formant frequencies (in Hz) of Barnaby phonating on different vowels.

6.5 A NEW METRIC TO COMPARE SINGER'S FORMANT CLUSTERS

In the last section, it could be observed that the Singer's Formant Cluster (SFC) exhibits many different patterns, and is difficult to compare them in practice: its constituent formants tend to vary in frequency, amplitude and bandwidth across singers, vowels and registers. To address this issue, a new metric of the SFC is suggested.

It consists in approximating the SFC by a Gaussian curve based on the amplitude and frequency of each of its N constituent formants. Each formant F_i is defined by its frequency $f(F_i)$ and its amplitude $A(F_i)$. Let us first define the centre, the amplitude and the width of the SFC:

1. The amplitude of the SFC is represented by Height and is defined as an average of the amplitude of each of its constituents:

$$\text{Height} = \frac{1}{N} \sum_{i=1}^N A(F_i) \quad (6.14)$$

2. The centre frequency of the SFC is represented by Centre and is defined as a weighted average of the frequency of its constitutive formants:

$$\text{Centre} = \frac{1}{N \cdot \text{Height}} \sum_{i=1}^N f(F_i) \cdot A(F_i) \quad (6.15)$$

A Gaussian curve centred at the origin is of the form

$$f(x) = Ae^{-\left(\frac{x}{a}\right)^2} \quad (6.16)$$

as plotted on Fig 6.61.

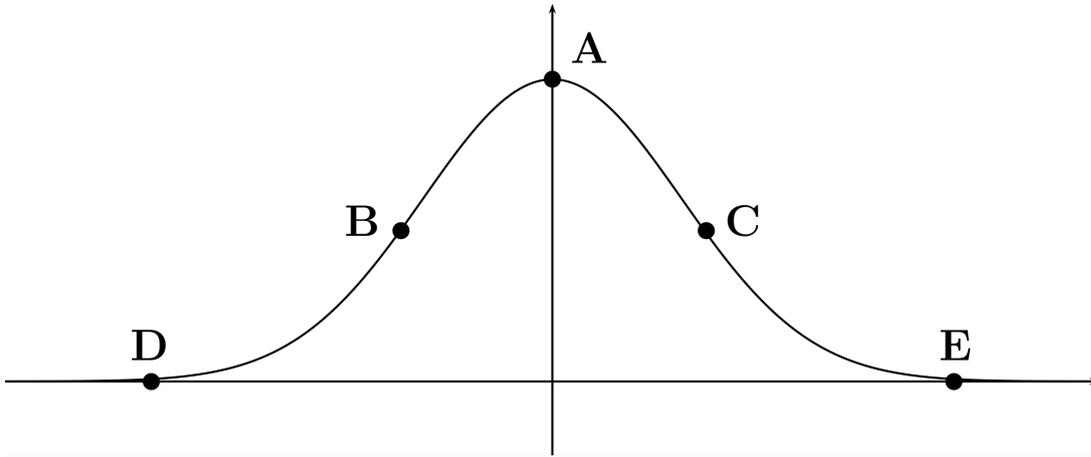


Figure 6.61: Gaussian function.

From the plot, it follows that A represents the Height of the SFC. Indeed,

$$\text{Height} = f(0) = Ae^{-\left(\frac{0}{a}\right)^2} = A \quad (6.17)$$

B and C are the inflexion points of the curve, i. e. the points where the curvature changes sign. It implies that the second derivative is null for these points:

$$\begin{aligned} [f(x)]''_{x=B,C} &= 0 \\ \left[-\frac{2Ax}{a^2} \cdot e^{-\left(\frac{x}{a}\right)^2} \right]'_{x=B,C} &= 0 \\ \left[-\frac{2A}{a^2} \cdot e^{-\left(\frac{x}{a}\right)^2} \cdot \left(-1 + \frac{2x^2}{a^2}\right) \right]_{x=B,C} &= 0 \end{aligned}$$

leading to

$$\begin{aligned} B &= -\frac{a}{\sqrt{2}} \\ C &= \frac{a}{\sqrt{2}} \end{aligned}$$

In statistics, the parameter a (a^2) represents the standard deviation (variance) of a normal distribution. To approximate the SFC, points B and C need to match the transfer function. It therefore defines two standard deviations, one on the left, one on the right, since the SFC is

not symmetric in general. The distance between B and C is called the Width of the SFC.

The graph of the Gaussian curve is then translated horizontally to the Centre and vertically to the offset given by the transfer function (see Fig 6.62).

The SFC Gaussian curve can then be written:

$$f = \text{Height} \cdot e^{-\left(\frac{x - \text{Centre}}{a}\right)^2} + \text{Offset} \quad (6.18)$$

where Offset is defined by

$$\text{Offset} = y_0 - \text{Height} \cdot e^{-\left(\frac{x_0 - \text{Centre}}{a}\right)^2} \quad (6.19)$$

with (x_0, y_0) being the coordinates of the spectral trough between F2 and F3. The parameter a is defined as

$$a = \frac{\text{Width}}{\sqrt{2}} \quad (6.20)$$

This parameter is declined with a left and a right version in case one wants to approximate the SFC by an asymmetrical Gaussian curve, as is the case here.

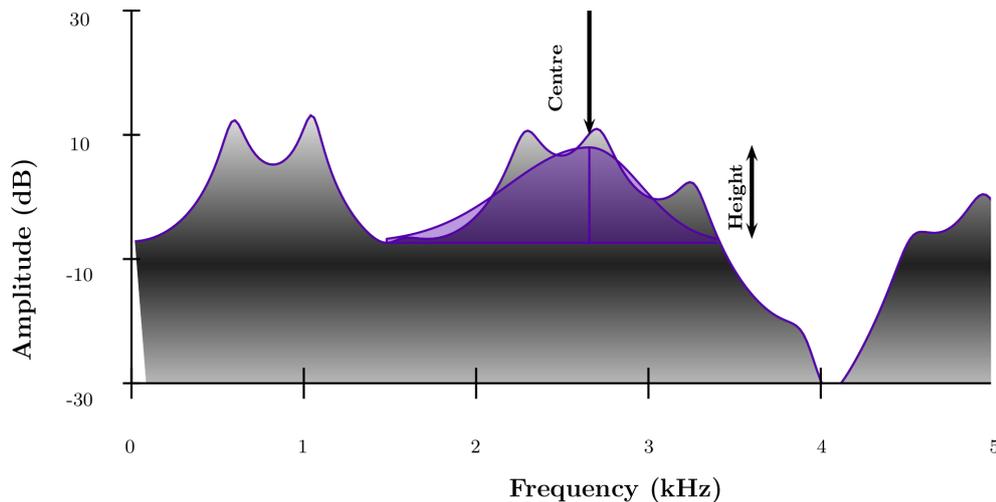


Figure 6.62: New metric for the SFC.

6.5.1 Test with prototypical cases

The behaviour of the metric is benchmarked with prototypical cases shown on Fig 6.63: a set of one to three peaks, each under the form of a Gaussian, is varied across its constitutive parameters such as its amplitude or height H and its bandwidth B .

For the convenience of computation, the figures have been drawn with an amplitude $H=6$ dB so that the bandwidth B (width at -3 dB) is measured at half the maximum amplitude.

Note that the parameter Width (6.20), referred to as the distance BC in Fig 6.61 measures the width of the function to be approximated (in black) at half its maximum height. For instance,

- In Fig **A**, $\text{Width}=B$;
- In Fig **B**, $\text{Width}=2\left(\frac{1}{2} \cdot B + \frac{1}{2} \cdot W\right) = W + B$;
- In Fig **C**, $\text{Width}=\frac{1}{2} \cdot B/2 + \frac{1}{2} \cdot W + \frac{1}{2} \cdot W + \frac{1}{2} \cdot B$.

One peak (A)

The SFC (in purple) suggested by the new metric fits exactly the Gaussian peak function, centred around zero.

Two peaks (B, C, D, E)

The metric is centred around zero for **B** and **C**. Indeed, the centre (6.15) of 2 peaks of the same height gives $\frac{1}{2}(-W/2 \cdot H + W/2 \cdot H) = 0$. The only minor difference between **B** and **C** is that the gaussian surface is totally symmetrical in **B** whereas it is slightly more prominent on the right side of **C**: the left Width of **C** is slightly smaller than its right counterpart. Both left and right Widths in **B** are equal to $\frac{1}{2} \cdot B + \frac{1}{2} \cdot W$ whereas the left Width of **C** is equal to $\frac{1}{2} \cdot B/2 + \frac{1}{2} \cdot W$ and the right Width to $\frac{1}{2} \cdot B + \frac{1}{2} \cdot W$.

D and **E** have a Height equal to $\frac{1}{2}(H/2 + H) = 3H/4$ and are centred around $\frac{1}{2 \cdot 3H/4}(-W/2 \cdot H/2 + W/2 \cdot H) = \frac{1}{2 \cdot 3H/4}WH/4 = W/6$. The

only difference between **D** and **E** is a slight asymmetry in **E** due to a left Width slightly smaller than the right Width.

Three peaks (F, G, H, I)

The metric is centred around zero for **F** and **G**. Indeed, the centre (6.15) of 3 peaks of the same height gives $\frac{1}{3}(-W \cdot H + 0 \cdot H + W \cdot H) = H$. The only minor difference between **F** and **G** is that the gaussian surface is totally symmetrical in **F** whereas it is slightly more prominent on the right side of **G**: the left Width of **G** is slightly smaller than its right counterpart.

H and **I** have a Height equal to $\frac{1}{3}(H/2 + H/2 + H) = 2H/3$ and are centred around $\frac{1}{3 \cdot 2H/3}(-W \cdot H/2 + 0 \cdot H/2 + W \cdot H) = \frac{1}{3 \cdot 2H/3}WH/2 = W/4$. There is no difference between **H** and **I** since the Width of both (width of the function at half its maximum height) equals $W + W + \frac{1}{2} \cdot B$.

When both the height and the bandwidth are changed, only the change in height matters (I).

To summarise:

- The Height and Centre of the SFC are only determined by the height and centre of its constituent peaks; the bandwidth has no influence.
- The Centre of the SFC is the average of the frequencies of its constituent peaks, weighed by their height.
- The Width of the SFC is predominantly determined by the distance between the external peaks and to a lesser extent by their respective bandwidth.

This new metric is aimed at facilitating the visual representation of complex patterns of SFC to compare them across singers/vowels/registers. Visually, it represents the statistical distribution of the peaks constituting the SFC.

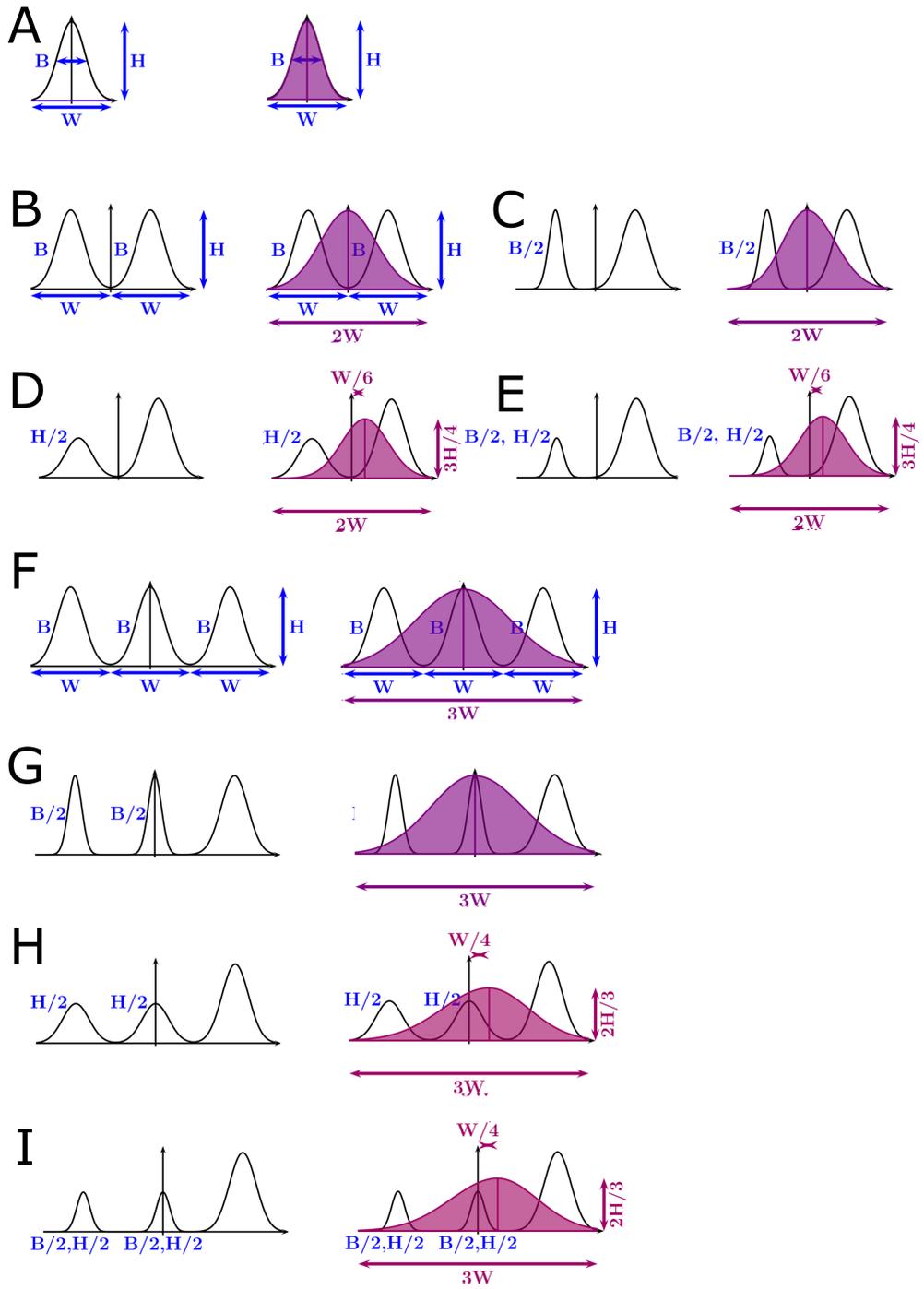


Figure 6.63: Benchmarking of the new metric against prototypical cases. The behaviour of the new metric is being scrutinised on prototypical cases made of 1, 2 or 3 peaks of a Gaussian form, whose parameters (height, width, bandwidth) are varied.

6.6 DATA VISUALISATION WITH THE NEW METRICS

The introduction of the new metric allows to compare the singer-s/vowels/registers from a different perspective.

6.6.1 Comparison with the dimensions

The SFC simulated by the new metric is compared with the SFC predicted by the Vocal Tract dimensions in Table 6.6 which shows successively: the anatomically predicted SFC, the metric SFC, and their relative difference expressed in %. For the dataset of six singers used for this thesis, the SFC predicted by the dimensions of the hypopharynx is consistent within 3% (2.52%) with the SFC obtained by the new metric: this adequation supports the formula (6.12) correlating the dimensions of the epilaryngeal tube, the pharyngeal width and the SFC.

SFC: predicted versus simulated

	SFC (pred) [Hz]	SFC (met) [Hz]	$\Delta(\%)$
Sophy	4,002	4,025	0.57
Maristela	3,529	3,574	1.28
Marisa	3,529	3,552	0.65
Timothy	2,971	3,046	2.52
Bartholomew	2,789	2,780	-0.32
Barnaby	2,671	2,652	-0.71

Table 6.6: Comparison between the SFC predicted (pred) from anatomical dimensions and the SFC obtained with the new metric (met).

6.6.2 Experimental versus Numerical

A comparison between the SFC obtained by the new metrics on numerical and experimental results shows good agreement between the experimental and numerical results (see Table 6.7)

SFC: simulation versus experimental

	SFC (num) [Hz]	SFC (exp) [Hz]	%
port	2,674	2,793	4.45
stern	2,627	2,603	-0.91
food	2,608	2,771	6.25

Table 6.7: SFC: comparison between the experimental and numerical results based on the new metric.

6.6.3 Singers Classification

The new metric of the SFC may visually facilitate the classification of singers into voice categories: having an arbitrary gradient scale for the SFC centre from 2 kHz (blue) to 4kHz (red), it clarifies the voice type which the singer belongs to. Fig 6.64 illustrates this principle on the six professional singers of this thesis, whose Vocal Tracts are represented on Fig . It is suggested that the use of the new metric helps visualising the differences across singers, from Fig 6.18, page 153 to Fig 6.64. On Fig 6.64, the SFC obtained with the new metric is superimposed to the transfer function of each singer. The SFC frequency centre determines the colour on the gradient scale, facilitating the identification of the voice type: towards red, the SFC and voice type are higher, towards blue, the SFC and voice type are lower. Note that Marisa and Maristela are quite close in terms of SFC, which is in agreement with the fact that they are both Mezzo-Sopranos.

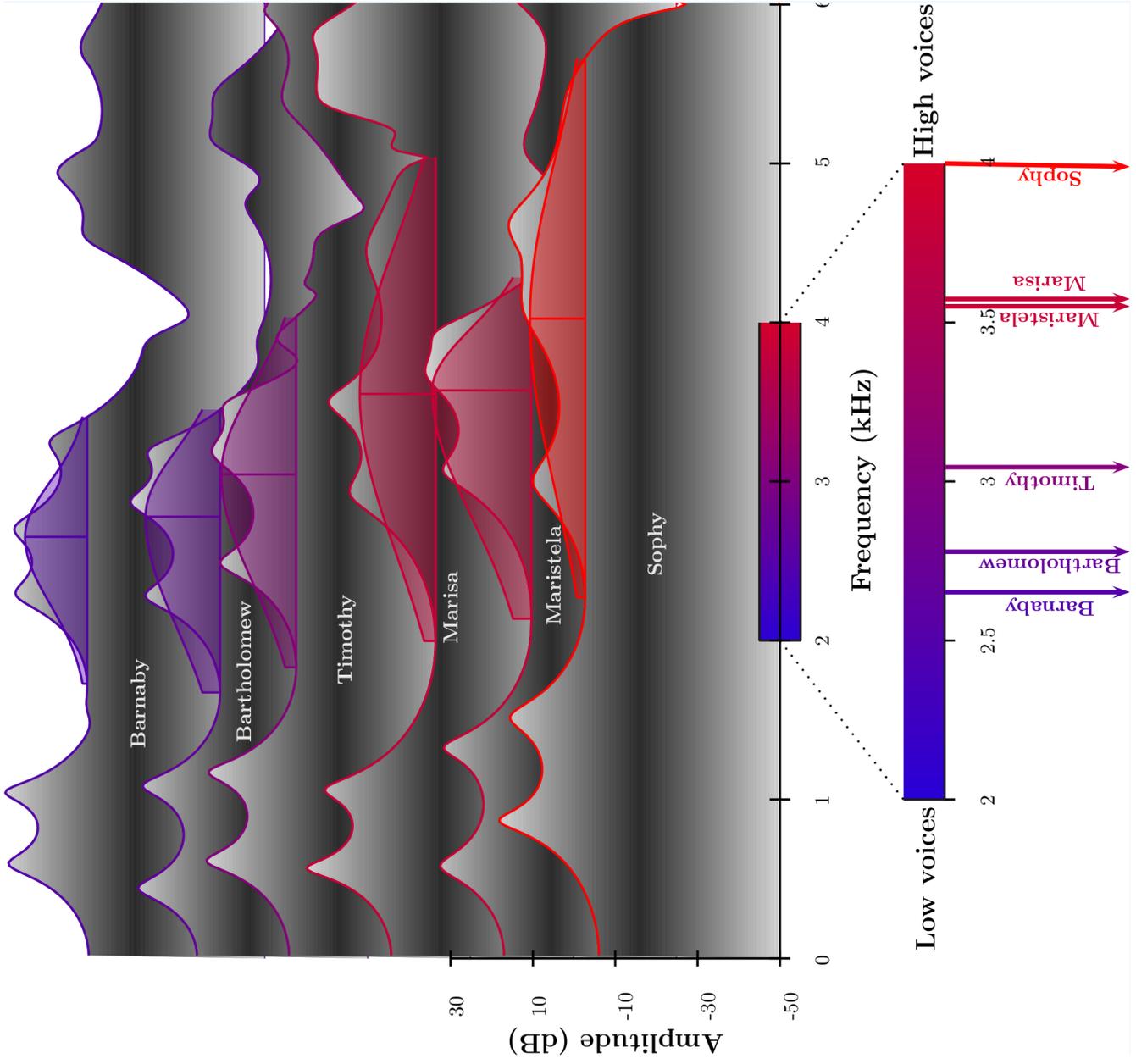


Figure 6.64: Changing voice type, from Bass-Baritone (Barnaby) to Soprano (Sophy) and the corresponding SFC obtained with the new metric.

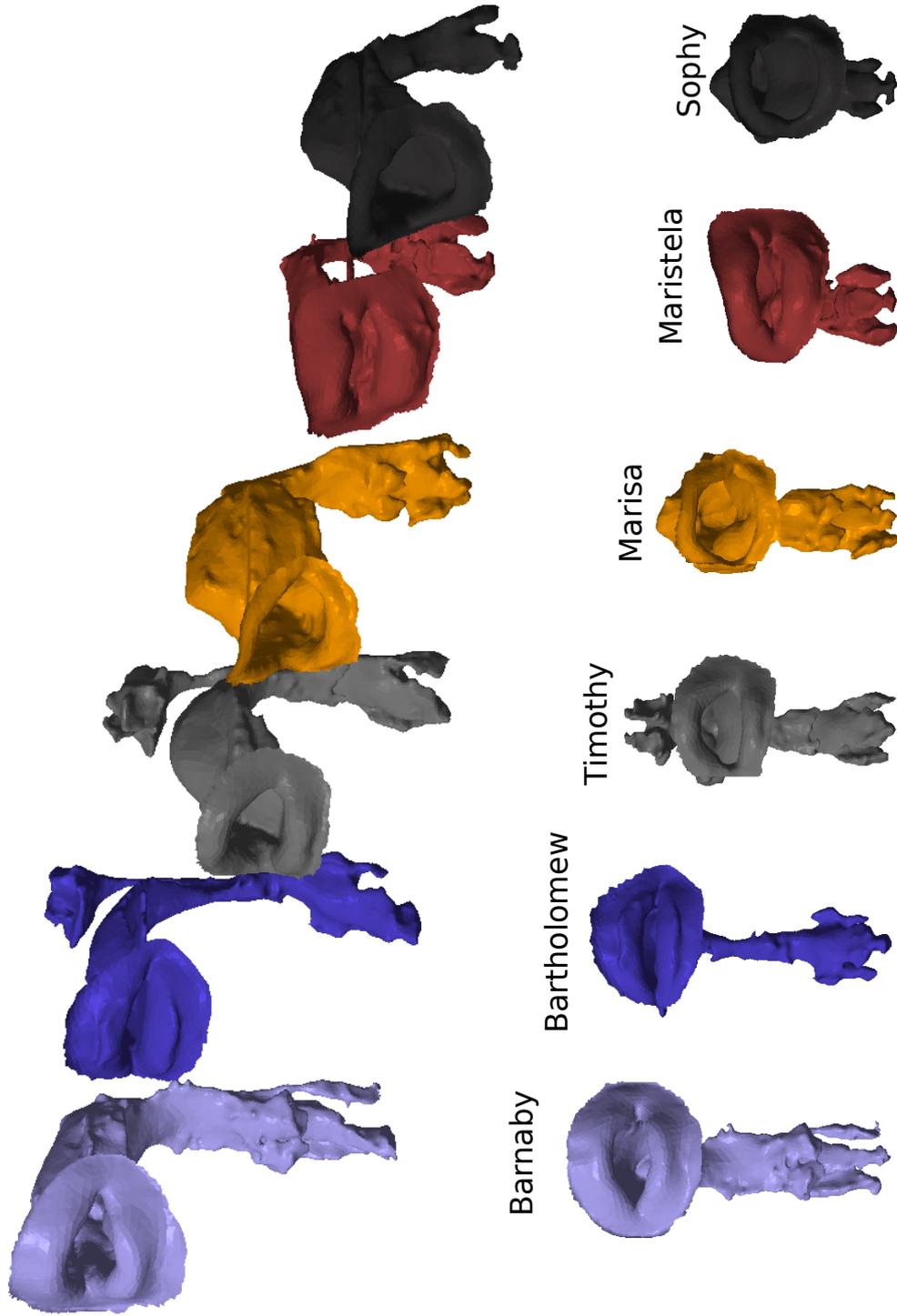


Figure 6.65: VTM-MRI of the six professional singers phonating on /hard/.

6.6.4 Same singer, different vowels

If the same principle is applied to the same singer, Barnabas, singing in different vowels (see Figs 6.66 and 6.67), it is observed that different SFC patterns appear, depending on the vowels being sung.

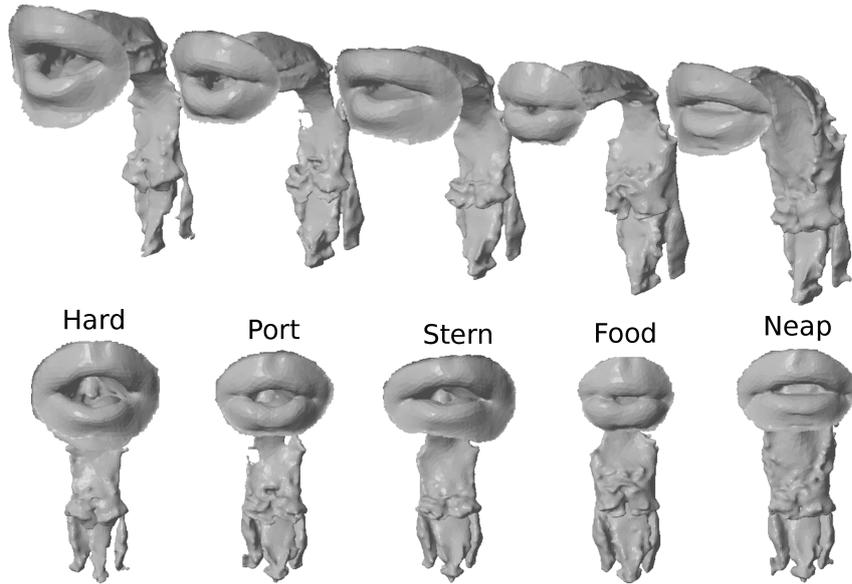


Figure 6.66: VTM-MRI of Barnaby singing on different vowels.

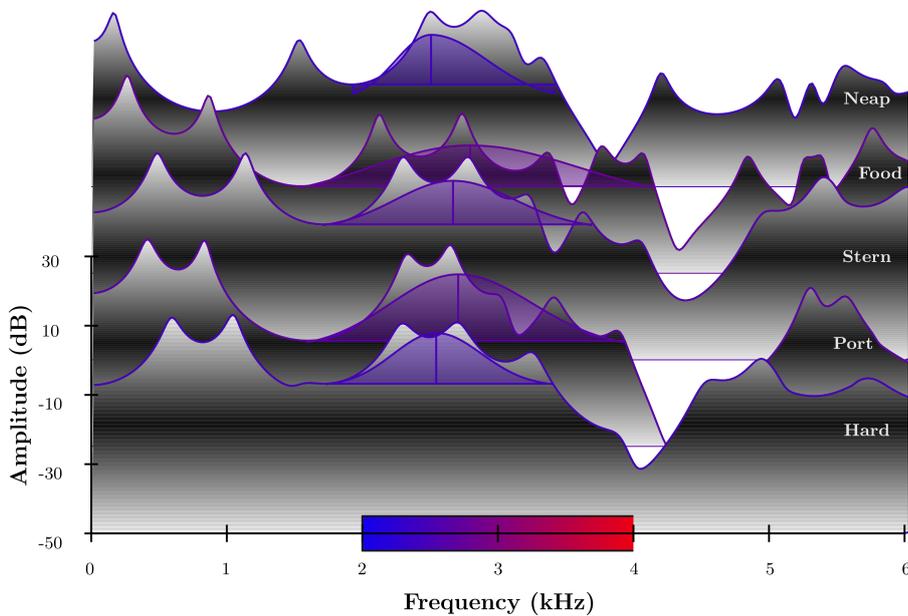


Figure 6.67: Barnaby singing on different vowels, with the SFC represented by the new metric.

6.6.5 Registers

Fig. 6.68 shows different SFC patterns when Bartholomew is singing on different registers. Fig 6.68 shows a strong nasal coupling via the velum except for the Falsetto register.

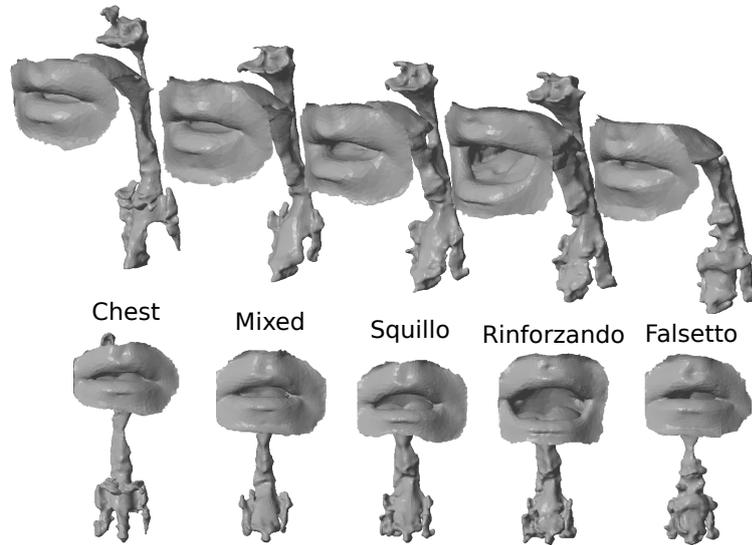


Figure 6.68: Bartholomew singing on different registers on /hard/.

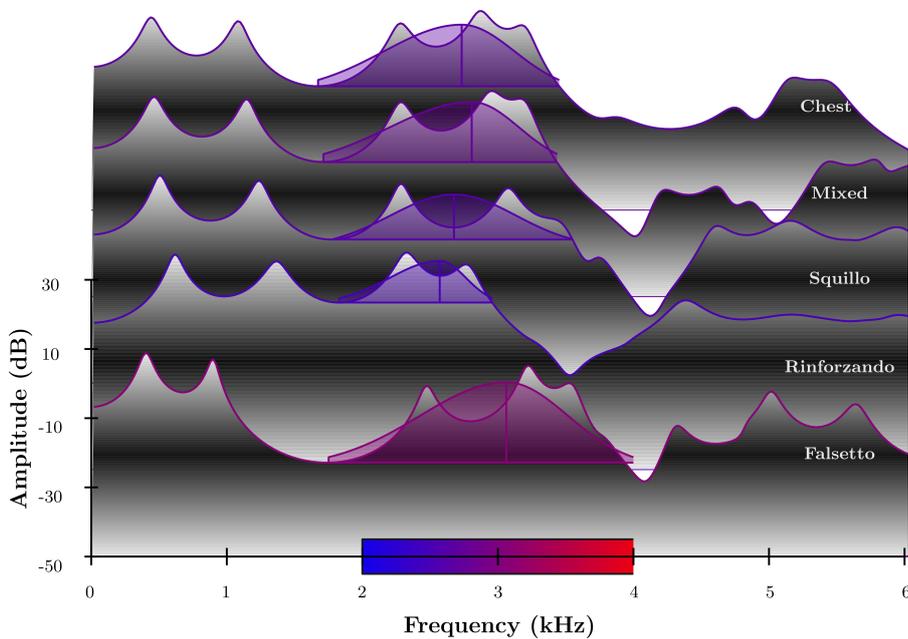


Figure 6.69: Bartholomew singing in different registers, with the SFC represented by the new metric.

6.6.6 Singer hallmark

Obtaining the SFC pattern for combinations vowel-register might give us more insight into the timbre characteristic of a singer's voice. An example is shown for Barnaby phonating on /hard/ in Fig 6.70. This figure shows a sum-up of the spectral effects of the hypopharyngeal cavities. On the upper left is the VTM-MRI-Barnaby-/hard/, with the epilarynx of Barnaby coloured in purple. On the upper right is the transfer functions of the VT (greyscale) and of the VTp (red). The purple Gaussian curve is the new metric aimed at facilitating the visual representation of the SFC. Note that its centre match closely the blue arrow which represents the SFC predicted by the anatomical dimensions of the hypopharynx (see formula (6.12) and Table 6.6). Note also that appending the epilarynx to the VTp spectrally shapes the VTp transfer function (in red) around the epilarynx resonance frequency (blue arrow) to form the VT transfer function (greyscale). On the bottom left, in red, the transfer function of the VTp, in greyscale, the transfer functions of the VT when the epilaryngeal tube is lengthened downwards by 2 mm steps from 0 mm to 10 mm. On the bottom right, in greyscale, the transfer function of the VT, in green, the transfer function of the VT without the piriform fossae. The green arrow is the antiresonance frequency generated by the piriform fossae.

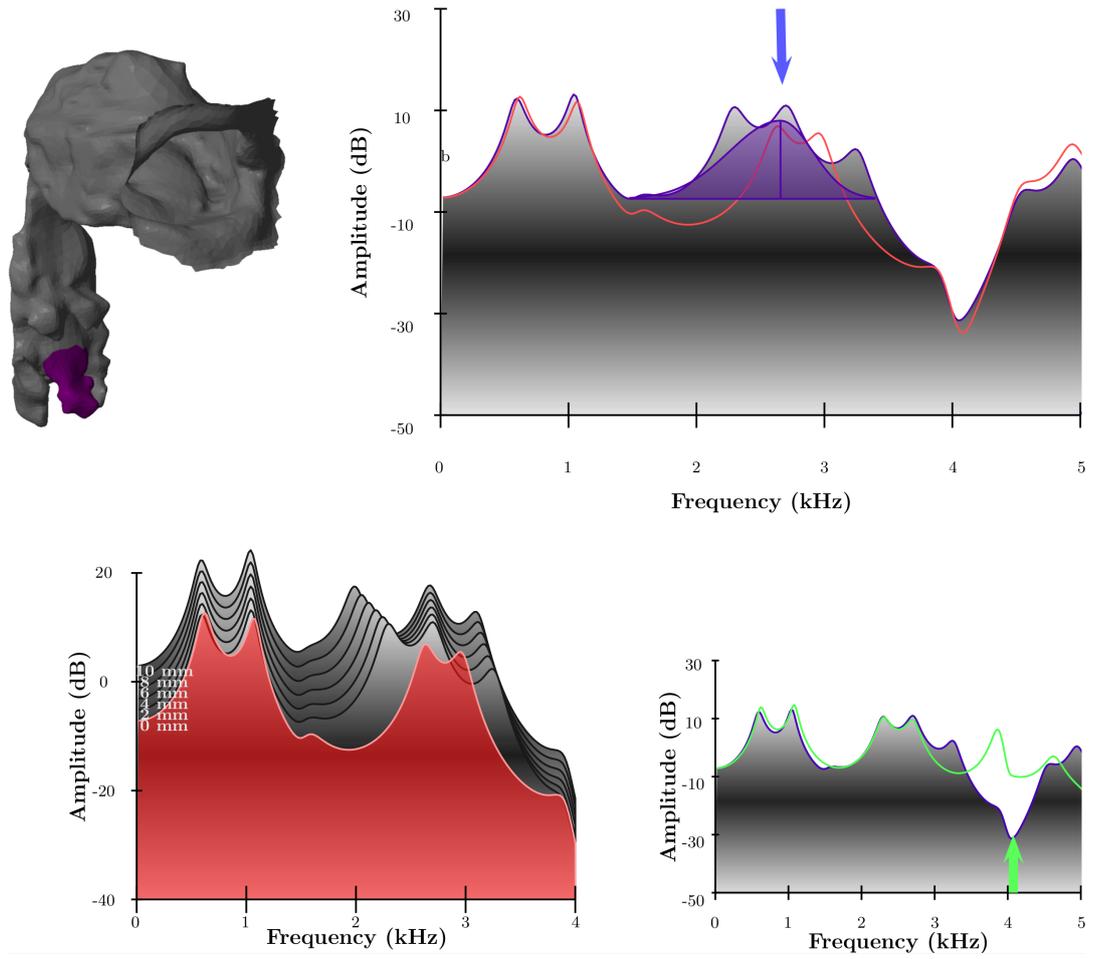


Figure 6.70: Barnaby singing on /hard/

Part IV

CONCLUSION

7 | CONCLUSION

Over the last few decades, many voice researchers have contributed to a better understanding of the human voice. Pioneers such as Fant, Sundberg, Titze and their predecessors unravelled the basic principles of voice production. Following their legacy, there are now several approaches to study voice production: the source-filter theory, which hypothesises no interaction between the source of the voice (glottal signal) and the filter (Vocal Tract), the glottal source itself, the Vocal tract itself, and the interactions between them. This thesis investigated the spectral impact of the hypopharyngeal cavities on the singing voice, within the framework of the source filter-theory and contributed in the following novel aspects:

- **A novel transducer-independent experimental method to measure the transfer function of 3D-printed Vocal Tract Models.**
- **Identification of the spectral impact of the epilarynx dimensions (length and radius) on Vocal Tract analogues (VTM-Ch&K).**
- **A new metric for the SFC is proposed, aimed at facilitating the visual comparison of the SFC patterns across singers, vowels and registers.**
- **A formula is suggested to predict the SFC frequency centre related to the dimensions of the hypopharynx.**

In the framework of this research, the findings:

- Support both numerically and experimentally the results found in the literature regarding the spectral impact of the piriform fossae and vallecule on the transfer function of MRI-based Vocal Tracts (VTM-MRI).

- Suggest, in the limitations of the data collected on six professional singers, that their SFC frequency centre is correlated to their hypopharyngeal dimensions and voice classification.

This chapter is structured as an overview of this thesis: the first part gave a background for the thesis, the second part introduced the methods (both numerical and experimental) used, the third part exposed the results. This is followed by a discussion, a model proposed for the hypopharyngeal cavities, the contributions, the shortcomings and further development.

7.1 BACKGROUND

This thesis has focused on the resonances of the filter (Vocal Tract) within the framework of the source-filter theory. A general knowledge of the production of speech/sound was first introduced, followed by a development of source-filter theory. Concentrating on the filter, the notions of impedance - resonance were explained and applied to the Vocal Tract. The formants, and especially the Singer's Formant Cluster (SFC) were explained. A literature review on the particular role of the hypopharyngeal cavities and their relationship to the SFC was given, but a consensus amongst the scientific community has not been reached regarding the spectral impact of the epilaryngeal tube. Furthermore, techniques to measure the Vocal Tract resonances were listed and explained. This was followed by a brief review on how previous researchers tried to relate the voice types of singers with parameters such as VT length, formant frequencies, SFC, VF lengths, etc.

7.2 BRIEF SUMMARY OF THE METHOD USED

The method chosen was based on the comparison between numerical (FEM, FVM) and experimental approaches to measure the resonances of VTMs (Vocal Tract Models).

The FEM (Finite Element Method) simulates the transfer function of the VTMs directly in the frequency domain whereas the FVM (Finite Volume Method) propagates an impulse in the VTMs, via the wave equation $\frac{\partial^2 \phi}{\partial t^2} = c^2 \nabla^2 \phi$ in the time domain. The impulse response is recorded and processed with an FFT (Fast Fourier Transform) to give the transfer function of the VTM. An absorption coefficient is implemented on the VTM walls and reflection-free boundary conditions and infinite elements are used at the radiation field boundary by the FVM and FEM respectively.

VTMs provide a series of shapes approximating the Vocal Tract, from the simplistic single tube (VTM-1), to a twin tube (VTM-2), a VTM based on Chiba & Kajiyama's Vocal Tract replicas (VTM-Ch&K, [2] from [14]) and eventually an MRI-based Vocal Tract of six professional singers while singing in a supine position (VTM-MRI). The different VTMs have been modelled and meshed for the use of the numerical methods. They have been 3D printed for the experimental approach. The MRI scan of the head of a singer while phonating on different vowels/registers gives slices in the 3 spatial directions, which need to be segmented to produce a 3D model of the vocal tract aimed at being meshed and used in the framework of numerical methods or being 3D-printed for experimental use.

7.2.1 A novel transducer-independent experimental method

The novel experimental approach exploits a technique widely used in room acoustics to measure an impulse response [34]: exciting the cavity of the VTM with an ESS (Exponential Sine Sweep) as an input signal, the response of which is convolved with the inverse filter of the ESS, splitting the linear impulse response from its harmonic distortions. The linear impulse response can then be converted into a transfer function by FFT. The adaptations/additions to the method used in [34] are as follows:

- **The acoustical response of a cavity (and not a room) is measured: the VTM is excited at the lips end and the sound output is recorded at the glottis end.**
- **The process is handled twice: once with the driver only, once with the driver+VTM, so that the transfer function of the standalone VTM is simply obtained by subtracting the two spectra (when the amplitude is represented logarithmically).**
- **A pre/post-envelope is applied on a frequency range (and not on a time interval) at the beginning and the end of the sweep, respectively.**

7.3 SUMMARY OF FINDINGS

The thesis proceeded to systematic acoustical analysis of a series of Vocal Tract models, by increasing order of complexity/realism: VTM-1, VTM-2, VTM-Ch&K and VTM-MRI. The first aspect developed concerns the spectral impact of the epilarynx, the extra resonance it generates and the spectral shaping around it. Its dimensions (length and radius) are varied and the spectral impacts are observed accordingly. These results are extended and being tested on real vocal tract shapes

(VTM-MRI). The effects of the piriform fossae and vallecula are investigated.

A theoretical background was first introduced to derive the resonances of a simple cylinder closed at one end and opened at the other. This represents the simplest model of a Vocal Tract (VTM-1) that is closed at the glottis end and opened at the lips end. The experimental result and numerical simulations matched the theoretically predicted transfer function.

Then, a smaller tube was appended at the glottis end of VTM-1, to form the VTM-2. The latter represents the epilaryngeal tube, whereas the former stands for the oropharyngeal tube. Introducing the OECC (Open End Correction Coefficient) which accounts for the small volume of air vibrating along with the air inside the VTM, the resonance frequencies of this twin tube are theoretically predicted and match those obtained by experimental set-up and numerical simulations. Appending a tube creates an extra resonance and spectrally shapes its neighbouring resonances. The centre frequency of this "spectral perturbation" corresponds to the first resonance frequency of the appended tube: the cross-section difference between the epilaryngeal tube and the oropharyngeal tube makes the former a $1/4$ -wave resonator. This supports the findings of [91, 93, 96, 96, 98, 108, 110, 27, 101, 50, 18, 57, 104, 105], see Fig 2.9 from [99] for an illustration of the generation of a singing formant (later termed the SFC). The method is applied for different lengths/radii of the epilaryngeal tube for VTM-2, VTM-Ch&K and VTM-MRI.

7.3.1 Identification of the spectral impact of the epilaryngeal tube on VTM-Ch&K

The new findings for VTM-Ch&K extend the results obtained for VTM-2:

- Decreasing the radius of the epilaryngeal tube increases the amplitude of the SFC, increasing the spectral power radiated around the SFC frequency centre (see Fig 6.10).
- No threshold ratio of 1:6 is observed as reported in [101]. A gradual increase in SFC prominence was observed instead.
- Increasing the length of the epilaryngeal tube shifts the centre of the SFC towards lower frequencies. This SFC shift is a key difference that is observed between singers when moving from a soprano to a bass range [16]; see Fig 2.14 from Dmitriev et al. [25].
- No significant variation of F_1 , F_2 (and to a lesser extent F_3) is in evidence when the epilaryngeal tube length is changed. Therefore, the vowel identification remains the same.

7.3.2 A new Metric for the SFC

Classically trained singers produce a broad energy peak in the region 2-4 kHz, termed the Singer's Formant Cluster, which enables them to be heard over an orchestra without amplification [119]. This SFC originates from the resonance of the epilaryngeal tube [101]. The spectra obtained for VTM-MRI showed that real singers exhibit different SFC patterns (see Fig 6.64). For the purpose of analysing and characterising the spectral differences across singers, vowels and registers, a new metric has been introduced to represent the SFC. It is aimed at facilitating the visual comparison between the SFC of different singers/vowels/pitches. It represents the statistical distribution

of the clustering resonance peaks contributing to the SFC: the peaks frequencies are averaged and weighted by their amplitude to yield the SFC frequency centre. The amplitude of the SFC is an average of the amplitude of its constituent peaks. The SFC is approximated by a Gaussian curve: this provides a straightforward visual comparison of different SFC patterns as well as the statistical distribution of the peaks constituting the SFC.

7.3.3 A formula to predict the SFC frequency centre based on the hypopharynx dimensions

From the findings of VTM-2 (see section 6.3.2) and VTM-Ch&K summarised in subsection 7.3.1, it appears that the SFC frequency centre is determined by two factors:

1. Predominantly by the length L of the epilaryngeal tube
2. To a lesser extent by the width of the pharyngeal opening PW and the cross-section A of the epilaryngeal tube at the pharyngeal opening (which accounts for the OECC for the interior of the neck of a resonator such as in (6.6), a fit formula from [51], valid as long as $\sqrt{A/\pi} < 0.2PW$)

This is illustrated in the suggested formula to predict the SFC frequency centre:

$$f = \frac{c}{4 \cdot (L + 0.48 \cdot (\sqrt{A} - \frac{2.5A\sqrt{\pi}}{\pi \cdot PW}))} \quad (7.1)$$

This suggests that, all other things being equal,

1. **A longer epilaryngeal tube generates a lower SFC centre**
2. **A wider epilaryngeal cross-section at the pharyngeal opening increases the acoustical length of the epilaryngeal tube, lowering the SFC centre, until $\sqrt{A/\pi} < 0.2PW$, so at the expense of the SFC amplitude.**

3. A wider pharyngeal opening increases the acoustical length of the epilaryngeal tube, lowering the SFC centre.

7.3.4 Findings

The hypothesis states that the hypopharyngeal cavities and the vallecula play a predominant spectral role in the timbre of the singing voice. The findings of this thesis suggest that:

- The amplitude of the SFC is inversely related to the cross-section of the epilaryngeal tube A at the pharyngeal opening: a smaller cross-section increases the amplitude of the SFC and therefore the acoustical power radiated in this frequency region, slightly changing the SFC centre frequency. This can be seen on Fig 6.10 for VTM-Ch&K, which builds on the results found for VTM-2 on Figs 6.6 and 6.7.
- The piriform fossae and epiglottic vallecula both act as side branches to the Vocal Tract, creating an anti-resonance in the 4-5 kHz region. It is likely that this trough perceptually enhances the effect of the SFC emerging from the spectrum, creating an acoustical energy gap which probably "isolates" the SFC signature in the acoustic output.
- The dataset of six professional singers provided in this thesis shows a clear correlation between the epilarynx dimensions, the SFC frequency centre and voice classification as can be seen on Fig 6.19: from Soprano to Bass-Baritone in the descending order, a lower voice type corresponds to a lower SFC and a longer epilarynx. This supports the findings of Dmitriev et al. [25] (Fig 2.14): they found that the dimensions of the VT were related to the production of singing formant and to the voice classification. The length of the epilaryngeal tube

(and both the width of the pharyngeal opening and the cross-section of the epilaryngeal tube to some extent, see formula 7.1) are therefore hypothesised to be related to the voice classification (see Fig 6.19): a longer epilarynx is related to a lower voice type.

- Based on the data of one singer in each case, the SFC of the same singer phonating on the same pitch on different vowels (Barnaby), or singing the same vowel on different registers (Bartholomew), exhibit different configurations of SFC (in terms of frequency centre, width and amplitude), see Figs 6.67 and 6.69 respectively.
- In the dataset obtained from six professional singers for this thesis, the SFC frequency centre predicted by the dimensions of the hypopharynx (formula 7.1) is consistent with the SFC frequency centre determined by the new metric within 3% (see Table 6.6).

7.4 DISCUSSION

A systematic acoustical analysis of Vocal Tract Models was conducted, by increasing complexity/realism order: VTM-1, VTM-2, VTM-Ch&K and VTM-MRI. The dataset of six professional singers provided in this study show a clear correlation between the SFC frequency centre and voice classification (Fig 6.19), supporting the results of the literature (Fig 2.14, from [25]). The predicted SFC frequency centre based on hypopharynx dimensions (formula 7.1) is consistent with the SFC frequency centre determined by the new metric within 3% (see Table 6.6) for the singers in this study. It is therefore suggested that the hypopharynx dimensions are related to the voice classification: the author hypothesises that the length of the epilarynx may determine the main voice classification while variations of the width of the pharynx-

geal opening might nuance this classification. Note that the oropharynx offers many different anatomical configurations across singers. The spectral action of the epilarynx on each of these oropharynges will therefore yields a great diversification of voices.

Titze et al. [108] states that *"a narrow epilarynx tube acts a bit like the mouthpiece of a brass instrument, matching the high internal impedance of the glottis to the lower impedance of the vocal tract and free space"*. As an impedance matcher, the epilarynx faces different impedance values of the oropharynx at the pharynx opening for different vowels. This might suggest why some vowels need to be adjusted and why some vowels-frequency range are more difficult to produce than others, depending on the singers.

Regarding vowel adjustments performed by singers at higher pitches, resonance tuning strategy [45] was observed for Sophy and Maristela. Note that both the tenor and the baritenor appeared to lower the velum (soft palate) which probably resulted in a strong acoustical coupling of the nasal tract and the oral tract, changing the amplitude and resonances of the oral tract, suggesting that this velar coupling was performed to tune one of the first two resonances to the fundamental frequency f_0 or one of its multiple.

7.5 MODEL FOR THE HYPOPHARYNGEAL CAVITIES

As an extension of the results of this thesis, a model for the spectral impact of the hypopharyngeal cavities on the voice output is proposed on the schematic of Fig 7.1:

The Vocal Tract proper determines the lower formants of the voice associated with its vowel quality. Appending the epilarynx, the piriform fossae and the vallecula reshapes the higher formants of the voice associated with its timbral quality.

The epilarynx length (plus the end correction depending on the cross-section of the epilarynx and the pharyngeal width) determines the frequency centre of the SFC, the epilaryngeal cross-section and the pharyngeal width impact on the amplitude and the width of the SFC. As a result, the VTp transfer function gains an extra resonance and its amplitude is locally increased around the first resonance frequency of the epilarynx (typically between 2-4 kHz). Appending the piriform fossae and the epiglottic vallecula adds side branches to the Vocal Tract, which create an anti-resonance around 4-5 kHz, spectrally enhancing the SFC even more.

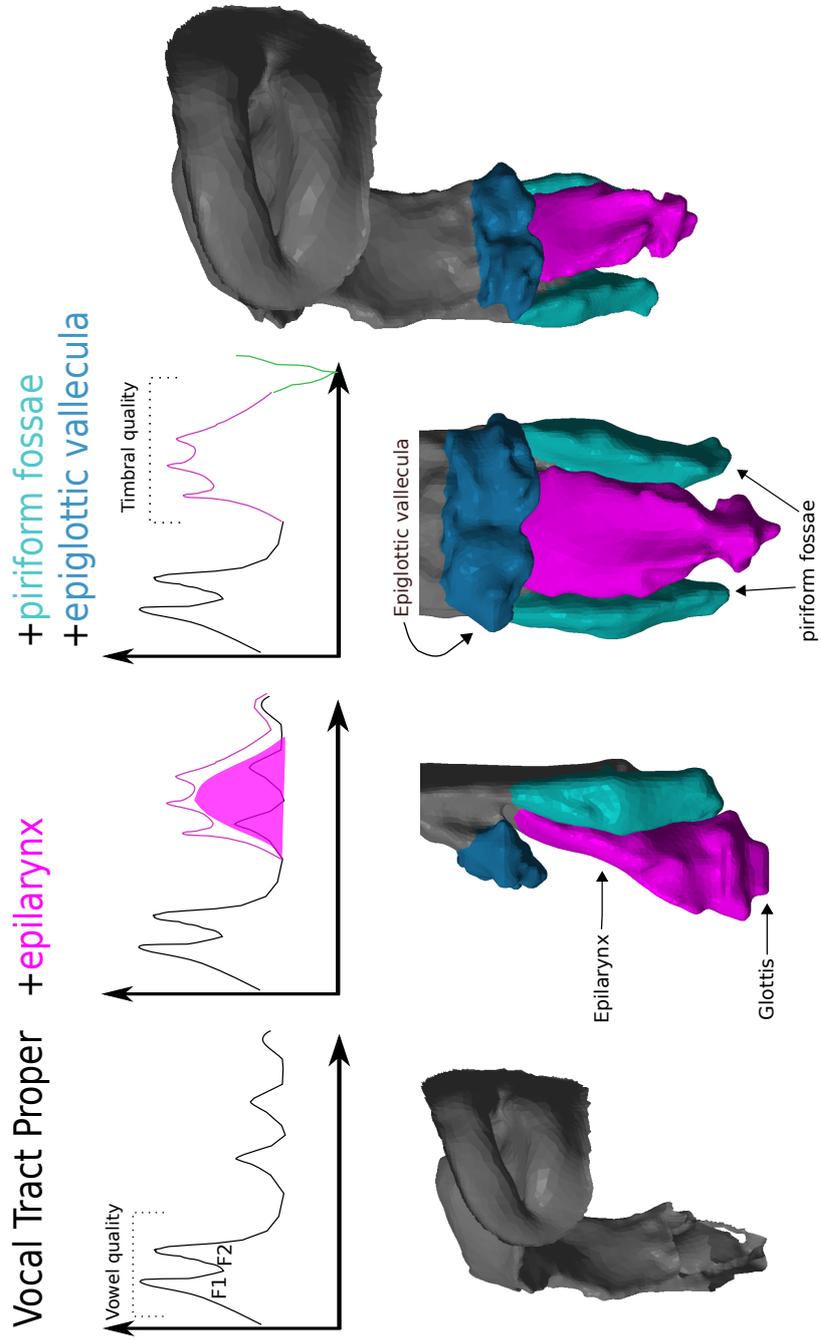


Figure 7.1: A schematic of the hypopharyngeal cavities model. From left to right, the transfer functions of the VTp, VTp+epilarynx, VTp+epilarynx+piriform fossae+epiglottic vallecula.

7.6 CONTRIBUTIONS

This thesis investigated the spectral impact of the hypopharyngeal cavities on the singing voice and contributed to these novel aspects:

- **A novel transducer-independent experimental method** aimed at measuring the transfer functions of cavities.
- **A new metric for the SFC** aimed at facilitating its visual representation and comparison across singers/vowels/registers.
- **A suggested formula to predict the SFC from the hypopharynx dimensions.**
- **The impact of the epilarynx dimensions on the transfer functions of VTM analogues (VTM-Ch&K).**
- **A suggested model of the hypopharyngeal cavities spectral impact.**

The findings of this thesis may be useful to the voice researchers community to highlight the potential relation between the hypopharynx structures and the voice classification. The new metric may help comparing more systematically the SFC patterns inter-singers and intra-singers (different vowels, registers).

7.7 SHORTCOMINGS

The methods used in this thesis show the following shortcomings:

1. The sample of professional singers was insufficient to draw general conclusions regarding voice classification.
2. The transfer functions, both measure and simulate the resonances of the VT + radiation at the lips, not only those of the

- VT. This results in the transfer function of the VT + the slope of the radiation.
3. This study was performed within the limits of the Source-Filter theory, so no interaction between the Vocal Folds and the Vocal Tract has been addressed.
 4. The accuracy of the MRI gave at best a 1 mm-resolution. This resolution could be slightly increased with a sinc ($\sin(x)/x$) interpolation performed by ITK-Snap between the MRI slices and a post-processing smoothing the surface with NetGen.
 5. The frequency resolution of the FEM is dependent on the size of the mesh.
 6. The FVM needs a timestep of the order of 10^{-9} s to perform a good time resolution with the same mesh as the FEM and results therefore in long simulations which would typically take about 6h. On the other hand, the frequency resolution is therefore largely increased.
 7. Only the oral tract transfer function was measured: no velar coupling was accounted for in the simulations.

7.8 FURTHER DEVELOPMENT

This thesis concludes with further developments which could potentially improve our understanding of the spectral impact of the hypopharyngeal cavities and their role in the timbral characteristics of the voice.

Regarding the numerical model, implementing a frequency-dependent admittance at the vocal tract walls, or even semi-porous walls could provide more accuracy in the amplitude of formants and their bandwidth. The location of the different formants, however, should not be

too affected, as seen in subsection 4.8.3. The implementation of the full set of 3D Navier-Stokes equations in the time domain, with the FVM would give a more accurate insight on the actual fluid dynamics at work in the VT, embodying turbulences, viscous layers, heat transfer and vorticity. In a model of voice production where the tract is coupled to the source (aero-acoustic framework), the use of Navier-Stokes would be relevant, but if the aim is only to investigate the resonances of the tract by propagating an impulse and measuring its response, the difference might not be relevant.

Concerning the measurement of the Vocal Tract geometry, an acoustic pharyngometer based on Acoustic Reflection Technology (ART) was used to measure the Vocal Tract length and volume of professional singers [120]. This works like a sonar: it sends an acoustic wave into the Vocal Tract and measures the reflections of this wave at each discontinuity, reconstructing the cross-sections and the volume. Further development of this technology might provide better volumetric representations of the Vocal Tract, which is less expensive than MRI. Optical wave propagation may also be considered as a non-invasive way to measure the tract geometry while phonating, and deduce the voice classification of a singer.

A more extensive study encompassing a greater number of professional singers would give more insight into the potential link between the hypopharynx dimensions, the SFC frequency centre and the voice classification. The new metric might provide a means to visually compare the SFC patterns of representatives of each voice category, in terms of vowels and registers.

Regarding the epilarynx, some questions remain unsolved and might be relevant to our understanding of voice classification. The epilarynx tube acts as an impedance matcher between the Glottis and the rest of the Vocal Tract [119]: it transmits optimally acoustical en-

ergy in several frequency bands downstream (to the tract) and upstream (sound waves from the tract on the Vocal Folds). The tessitura and the voice classification are then probably linked through the epilarynx, which acts as an impedance matcher, facilitating the vibration of the folds at several frequencies and enhancing the Vocal Tract resonance in some frequency ranges, and conversely. However, it is not yet understood whether the length of the Vocal Folds, and therefore the possible range of f_0 determines the length of the epilarynx or the other way round. Potentially, both probably evolve in such a way that for a healthy voice, the first resonator of the VT, i. e. the epilarynx and the glottal signal source, i. e. the Vocal Folds match each other in terms of impedance. However, more research is needed to investigate these aspects.

To what extent can the epilarynx tube be extended or constricted (changing the SFC centre, amplitude and width)? Can it be trained? What are the geometric modifications of the epilarynx at work in the different registers? In a singer's ease to produce a musically appealing and healthy operatic sound, in what proportion is it due to the epilarynx/oropharynx innate geometry or to the training to optimise the transmission of acoustical energy with the morphology the singer has been provided with?

Part V

APPENDIX

A | ACOUSTIC MODES FOR AN OPEN CYLINDER

The eigen-function $\Psi(\mathbf{r}, n)$ satisfies the Helmholtz equation at any point $\mathbf{r}(r, \phi, z)$ in the cylinder

$$[\nabla^2 + k_0^2(n)]\Psi(\mathbf{r}, n) = 0 \quad (\text{A.1})$$

where $k_0(n)$ are the eigenvalues.

The eigenfunctions must satisfy the rigid boundary condition at the surface of the open cylinder, i.e.

$$\frac{\partial \Psi(\mathbf{r}, n)}{\partial \hat{\mathbf{n}}} = 0 \quad (\text{A.2})$$

where $\hat{\mathbf{n}}$ is the surface normal.

In cylindrical coordinate system, ∇^2 is given by

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \quad (\text{A.3})$$

To solve the eigen-value problem, we assume the eigen-functions can be factored into a form $\Psi(\mathbf{r}, n) = R(r)e^{im\phi}Z(z)$. The Helmholtz equation is reduced to

$$\frac{1}{R} \left[\frac{\partial^2 R(r)}{\partial r^2} + \frac{1}{r} \frac{\partial R(r)}{\partial r} \right] - \frac{m^2}{r^2} + \frac{1}{Z(z)} \frac{\partial^2 Z(z)}{\partial z^2} + k_0^2 = 0 \quad (\text{A.4})$$

Applying separation of variables, we obtain independent equations for the axial factor $Z(z)$ and the radial factor $R(r)$. The axial factor $Z(z)$ satisfies

$$\frac{\partial^2 Z(z)}{\partial z^2} + k_z^2 Z(z) = 0 \quad (\text{A.5})$$

and also boundary conditions

$$\begin{cases} \frac{\partial Z(z)}{\partial z} = 0 & \text{at the closed end } z = -L \\ Z(z) = 0 & \text{at the open end } z = 0 \end{cases} \quad (\text{A.6})$$

Therefore, the axial factor $Z(z)$ is the solution of the above equations, which is $Z(z) = \sin(k_z z)$, and the eigen values $k_z = \frac{2n_z+1}{L} \frac{\pi}{2}$ for $n_z = 0, 1, 2, 3, \dots$

The radial factor must satisfy

$$\frac{\partial^2 R(r)}{\partial r^2} + \frac{1}{r} \frac{\partial R(r)}{\partial r} - \left(k_r^2 - \frac{m^2}{r^2} \right) R(r) = 0 \quad (\text{A.7})$$

where $k_r^2 = k_0^2 - k_z^2$, and the rigid boundary condition is

$$\frac{\partial R}{\partial r} = 0, \quad r = a \quad (\text{A.8})$$

The solution to the above equation is the cylindrical Bessel function $J_m(k_r r)$, where m is the order and k_r the eigenvalue. The eigenvalues ($k_{r,m}(n)$) are the roots of the derivative of the cylindrical Bessel function at $r = a$, that is

$$J'_m(k_{r,m}(n)a) = 0 \quad (\text{A.9})$$

Therefore, the eigenfunction for an open rigid cylinder is given by

$$\Psi_{n_z, m}(n) = J_m(k_{r,m}(n)r) e^{im\phi} \sin(k_z z) \quad (\text{A.10})$$

B | FINITE VOLUME METHOD

The **Finite Volume Method** is a discretisation method for the approximation of partial differential equations translating the conservation or balance of one or more quantities [29]. This method suits a wide range of numerical problems such as heat and mass transfer, fluid dynamics or petroleum engineering [30]. The partial differential equations (PDEs) are called *conservation laws* and typically relate the partial derivatives of unknowns such as density, temperature, pressure, with the variables of the domain under investigation (time, space, ...).

As in the FEM, the domain is partitioned, or *meshed* into smaller cells called *control volumes*. Each control volume is given a local balance equation, the result of the integration of the PDEs. This set of equations is then discretised with respect to a set of discrete unknowns [29]. The critical point in FVM resides in the fact that the numerical fluxes at the boundaries between adjacent cells must be:

- **conservative**, i. e. the flux coming from the control volume A to the control volume B must be the opposite of the one from B to A at any time.
- **consistent**, i. e. as the mesh size vanishes, the numerical flux tends to the continuous flux.

This results in a system of discrete equations with discrete unknowns which are related to one another either linearly or non-linearly. The system is then solved or approximated, using for instance iterative solvers and Newton methods in the case of linear and non-linear equations respectively.

B.1 BASIC PRINCIPLES

The FVM is used for the discretisation of conservation laws. Let $q(\mathbf{x}, t)$ denote a quantity depending on time (t) and space coordinates (\mathbf{x}). A conservation law expresses the conservation of a quantity q , e. g. the mass, the energy, the momentum, the number of moles of a chemical reactant, etc. A local form of the conservation law can be expressed as follows:

$$\boxed{q_t(\mathbf{x}, t) + \nabla \cdot \mathbf{F}(\mathbf{x}, t) = f(\mathbf{x}, t)} \quad (\text{B.1})$$

where

- q_t is the partial time derivative of q
- $\nabla \cdot$ is the space divergence operator: $\nabla \cdot \mathbf{F} = \frac{\partial F_1}{\partial x_1} + \dots + \frac{\partial F_d}{\partial x_d}$ with $\mathbf{F} = (F_1, \dots, F_d)$
- \mathbf{F} is the flux: it represents the transport mechanism of q
- f is the source/sink term: it defines any volumetric exchange/-force, e. g. product during a chemical reaction, gravity, etc.

1D EULER EQUATION

As an example, here is the 1D Euler Equation for real gases at equilibrium [30]. With

$$q = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix} \quad \mathbf{F} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E + p) \end{pmatrix}$$

where ρ is the density, u the velocity, E the total energy per unit volume and p the pressure.

The system of equations contains more unknowns than equations. It can be closed if p and E are related to the specific volume $\tau = \frac{1}{\rho}$ and the entropy s , yielding to 2 constitutive laws:

$$p = \frac{\partial \epsilon}{\partial \tau}(\tau, s), \quad E = \rho \left(\epsilon(\tau, s) + \frac{u^2}{2} \right)$$

where ϵ is the internal energy per unit mass, which is a given function of τ and s .

The equation (B.1) can be understood as the expression of the local conservation of q in an infinitesimal domain: it is equivalent to

$$\int_K q(\mathbf{x}, t_2) d\mathbf{x} - \int_K q(\mathbf{x}, t_1) d\mathbf{x} + \int_{t_1}^{t_2} \int_{\partial K} \mathbf{F}(\mathbf{x}, t) \cdot \mathbf{n}_K(\mathbf{x}) d\gamma(\mathbf{x}) dt = \int_{t_1}^{t_2} \int_{\partial K} f(\mathbf{x}, t) d\mathbf{x} dt \quad (\text{B.2})$$

for any subdomain K and for any time t_1 and t_2 , where \mathbf{n}_K is the unit vector normal to the boundary ∂K , at the point \mathbf{x} , pointing outward K . (B.2) expresses the conservation of the quantity q in the cell K between the times t_1 and t_2 . $d\mathbf{x}$ represents the integration symbol for the d -dimensional Lebesgue measure in \mathbb{R}^d and $d\gamma$ is the integration symbol for the $(d - 1)$ -dimensional measure [30].

B.2 TIME DISCRETISATION

Let $(t_n)_{n \in \mathbb{N}}$ denote an increasing time series with $t_0 = 0$. The time step can be variable or constant (as in this demonstration). Let $k \in \mathbb{R}_+^*$ denote the time step so that $t_n = nk$ for $n \in \mathbb{N}$. Note that B.1 can also be written with a time-space divergence. This leads to two possibilities in terms of time discretising (see Table B.1):

Implicit Scheme	Explicit Scheme
TIME-SPACE DIVERGENCE	SPACE DIVERGENCE
<i>space-time</i> finite volume discretisation	<i>space</i> finite volume discretisation + time finite difference scheme
Conservation Law integrated over time and space	Conservation Law integrated over space Time derivative approximated by a time difference scheme

Table B.1: Comparison between implicit and explicit schemes.

This classifies the time discretisation in two families of schemes: implicit and explicit. Explicit Euler, for instance, discretises the time derivative as such:

$$(q)_t \cong \frac{q^{(n+1)} - q^{(n)}}{\kappa}$$

with $q^{(n+1)}$ and $q^{(n)}$ representing the quantity q at times $n + 1$ and n respectively. Note that implicit schemes and higher orders might also be used.

B.3 SPACE DISCRETISATION

For a space discretisation of (B.1), a mesh \mathcal{T} of the domain Ω of \mathbb{R}^d needs to be introduced, such that $\overline{\Omega} = \cup_{K \in \mathcal{T}} \overline{K}$, where K is an element of \mathcal{T} , an open subset of Ω called a control volume [30].

$q_K^{(n)}$ represents an approximation of the quantity q in the volume control K at the time t_n . The principle of the finite volume method is

to integrate (B.1) over each cell K of the mesh \mathcal{T} to obtain the conservation under a nonlocal form for the volume K . Using Euler explicit, this gives [30]:

$$\int_K \frac{q^{n+1}(\mathbf{x}) - q^n(\mathbf{x})}{k} d\mathbf{x} + \int_{\partial K} \mathbf{F}(\mathbf{x}, t_n) \cdot \mathbf{n}_K(\mathbf{x}) d\gamma(\mathbf{x}) = \int_{\partial K} f(\mathbf{x}, t_n) d\mathbf{x} \quad (\text{B.3})$$

The last step is to approximate the flux, i. e. to express $\mathbf{F}(\mathbf{x}, t_n) \cdot \mathbf{n}_K(\mathbf{x})$ across the boundary ∂K of each control volume (see Fig B.1), in terms of $q_L^{(n)}$, $L \in \mathcal{T}$. Let $K|L = \bar{K} \cap \bar{L}$ with $K, L \in \mathcal{T}$, the exchange term (from K to L)

$$\int_{K|L} \mathbf{F}(\mathbf{x}, t_n) \cdot \mathbf{n}_K(\mathbf{x}) d\gamma(\mathbf{x})$$

from the control volume K to the control volume L between times t_n and t_{n+1} , is approximated by some quantity $\mathbf{F}_{K,L}^{(n)}$, which is a function of $u_M^{(n)}$, $M \in \mathcal{T}$. Note that $\mathbf{F}_{K,L}^{(n)} = 0$ if the Hausdorff dimension of $\bar{K} \cap \bar{L}$ is less than $d - 1$ (e. g. $\bar{K} \cap \bar{L}$ is a point in the case $d = 2$ or a line in the case $d = 3$)[30]. Here are two important features of the finite volume method:

1. Conservativity: $\mathbf{F}_{K,L}^{(n)} = -\mathbf{F}_{L,K}^{(n)} \quad \forall K, L \in \mathcal{T} \text{ and } \forall n \in \mathbb{N}$
2. Consistency of the approximation of $\mathbf{F}(\mathbf{x}, t_n) \cdot \mathbf{n}_K(\mathbf{x})$, which needs to be defined for each relation between \mathbf{F} and the unknowns.

B.4 COMPARISON WITH OTHER METHODS

The Finite Volume Method is often compared with the Finite Difference Method (FDM) and the Finite Element Method (FEM), and although there are similarities, the main differences are described below [30].

The Finite Difference Method is based on a set of discretisation points, each of which is assigned one unknown and one discretised

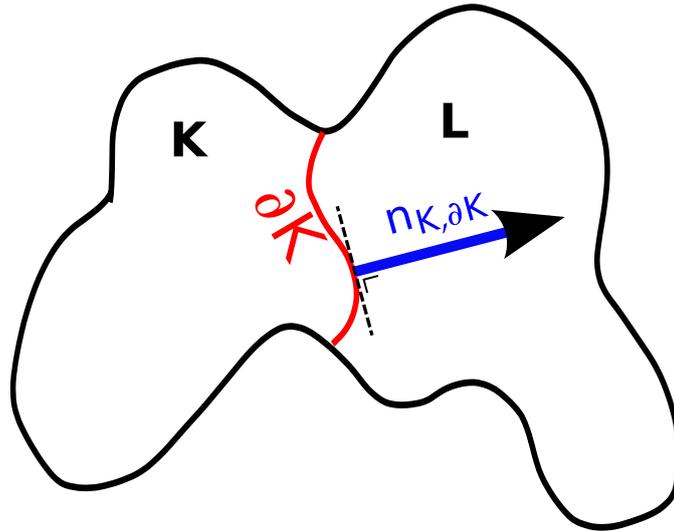


Figure B.1: Boundary between two adjacent volume controls K and L.

equation [30]. At each point, the derivatives are replaced by a finite difference, using a Taylor expansion to the neighbouring points. In comparison with the FDM, the FVM discretises the balance equation and not the PDE itself. One of the shortcomings of FDM is that it cannot deal with discontinuous coefficients, for instance in the case of heterogeneous media. By opposition, with the FVM, this problem can be handled quite easily by choosing the control volumes adequately so that their boundaries coincide with the separation between the different media.

The Finite Element Method is based on a variational formulation, for both the discrete and the continuous problems. It consists of multiplying the original equation by a so-called *test function*. The continuous unknown is approximated by a linear combination of *shape functions*, which are the *test functions* for the discrete variational formulation (this is called the *Galerkin Expansion*) [30]. The resulting equation is then integrated over the whole domain. Sometimes, the FVM is called as a discontinuous FEM, in the sense that you can choose a test function defined as $1_K(\mathbf{x}) = 1$, if $\mathbf{x} \in K$, and $1_K(\mathbf{x}) = 0$, if $\mathbf{x} \in \bar{K}$ and use a linear combination of shape functions for the discrete unknown.

C | FINITE ELEMENT METHOD

FEM is a very successful method across the engineering world, mostly because of two principles on which it is based [75]:

1. The domain Ω in \mathbb{R}^d is partitioned into N small, non-overlapping domains, called *finite elements*, over which functions are approximated by local functions, usually polynomials [42].
2. The boundary- and initial-value problems are formulated in a *weak form*, or integral form, so that each subdomain contributes to the whole domain by summing up the integrals over each domain.

The second point probably played an important role in the major success encountered by FEM, because the local subdomain approximations are summed up to determine the contributions to the full domain.

C.1 BASIC PRINCIPLES

The Finite Element Method consists in seeking an *approximated* function $\mathbf{u}^h(\mathbf{x}, t)$ of the exact solution $\mathbf{u}(\mathbf{x}, t)$, continuous piecewise on subdomains of the domain Ω . The N subdomains are defined such as

$$\bigcup_{i=1}^N \Omega_i = \Omega \quad \text{and} \quad \overline{\Omega_i} \cap \overline{\Omega_j} = \emptyset \quad \forall i \neq j \quad (\text{C.1})$$

where $\overline{\Omega_i}$ designates the interior of Ω_i . In other words, $\{\Omega_i\}$ is a partition of Ω .

Note that $\mathbf{u}(\mathbf{x}, t)$ can be a scalar, vectorial or tensorial field, dependent on the space coordinates in the domain Ω and the time t .

The fields $\mathbf{u}_i^h(\mathbf{x}, t)$, defined for each sub-domain, are chosen amongst an arbitrary family of functions (which are generally polynomials in the FEM).

- The local field family is called *element interpolation functions space*
- The global field family is called *domain interpolation functions space*

The field in each subdomain Ω_i is determined by a finite number of values of the field (or its derivatives) evaluated at points of the subdomain arbitrarily chosen called *nodes*. The local field is then an interpolation between the values at the nodes. The subdomain together with its interpolation is called *element* *.

Seeking a solution in the sense of FEM consists in determining which local field is attributed to subdomains so that the global field $\mathbf{u}^h(\mathbf{x}, t)$ obtained by juxtaposition of these local fields $\mathbf{u}_i^h(\mathbf{x}, t)$ gets close to the exact solution $\mathbf{u}(\mathbf{x}, t)$.

The quality of the approximation depends on:

- The partition of the domain: refinement of the mesh, adaptable mesh to the geometry, structured/unstructured grid, etc.
- The choice of the family of local fields: polynomials, etc.
- The continuity conditions imposed at the boundaries inter-subdomain:
 C_0, C_1, \dots

To solve a problem with the FEM, the following steps would be employed [42]:

1. Formulate a problem in physics as a system of (partial) differential equations to be satisfied in every point of a domain Ω , with boundary conditions on the boundary $\partial\Omega$
2. Derive an integral formulation of the differential system to be solved, along with its boundary conditions: this is the *variational formulation of the problem*
3. Divide Ω in subdomains: this is the *meshing*
4. Choose the the family of local fields, i. e. :
 - the position of the nodes in the subdomains
 - the polynomials (or other functions) which define the local field with respect to the values (and possibly its derivatives) at the nodes
5. Formulate the problem discretely: this is the *discretisation*. Indeed, each approximated solution is entirely determined by its value at the nodes of the elements. It then "only" needs to find the values to confer to the nodes to describe an approximated solution. The fundamental principle of the FEM can be summed up in two questions:
 - a) How to choose the discrete problem whose solution is "close" to the exact solution?
 - b) How to define "close"?
6. Solve the discrete problem: this is the *resolution*

These principles are better understood when illustrated by a simple example. To that purpose, the case of Poisson equation in 2D will be examined, with first an incursion to the Garlekin method to explain what the variational form is about.

C.2 GARLEKIN METHOD

The purpose is to find an approximated solution of a problem *. The variational form can be written under the form:

$$\text{Find } u \text{ in } V \text{ such as } \quad a(u, v) = \langle f, v \rangle \quad \forall v \in V \quad (\text{C.2})$$

where

- V is vectorial space whose elements are functions
- $(u, v) \mapsto a(u, v)$ is a bilinear function on $V \times V$
- $v \mapsto \langle f, v \rangle$ is a linear form on V

In the case of Poisson problem with homogeneous Dirichlet conditions, the vectorial space on which we seek solutions is $V = \{u \text{ continuous and } C_1 \text{ piecewise on } \Omega, u = 0 \text{ on } \partial\Omega\}$ and the forms $a(., .)$ and $\langle f, . \rangle$ are defined by

$$a(u, v) = \int_{\Omega} u'(x)v'(x) dx, \quad \langle f, v \rangle = \int_{\Omega} f(x)v(x)dx \quad (\text{C.3})$$

The solution space V being of infinite dimension, the principle of Garlekin method consists of replacing it by a vectorial space V^h of finite dimension and solving the approximated problem:

$$\text{Find } u^h \text{ in } V^h \text{ such as } \quad a(u^h, v^h) = \langle f, v^h \rangle \quad \forall v^h \in V^h \quad (\text{C.4})$$

As V_h is of finite dimension, solving (C.4) consists of solving a linear system (i. e. inverting a matrix). Indeed, if (ϕ_1, \dots, ϕ_N) is a base of V^h , the approximated solution u^h can be decomposed onto that base following

$$u^h(x) = \sum_{j=1}^N u_j \phi_j(x) \quad (\text{C.5})$$

Solving (C.4) is then equivalent to finding a vector $\mathbf{U} = (u_1, \dots, u_N) \in \mathbb{R}^N$ such as

$$\begin{aligned} & a \left(\sum_{j=1}^N u_j \phi_j(x), v^h \right) = \langle f, v^h \rangle, \quad \forall v^h \in V^h \\ \Leftrightarrow & \sum_{j=1}^N a(\phi_j(x), v^h) u_j = \langle f, v^h \rangle, \quad \forall v^h \in V^h \\ \Leftrightarrow & \sum_{j=1}^N a(\phi_j(x), \phi_i) u_j = \langle f, \phi_i \rangle, \quad \forall i \in \{1, \dots, N\} \end{aligned}$$

since (ϕ_1, \dots, ϕ_N) is a base of V_h . By setting the matrix

$$\mathcal{A} = (\mathcal{A}_{ij})_{1 \leq i, j \leq N} \in \mathbb{R}^{N \times N}, \quad \mathcal{A}_{ij} = a(\phi_j, \phi_i) \quad (\text{C.6})$$

and the vector

$$\mathbf{F} = (F_i)_{1 \leq i \leq N} \in \mathbb{R}^N, \quad F_i = \langle f, \phi_i \rangle \quad (\text{C.7})$$

we obtain the linear system

$$\mathcal{A}\mathbf{U} = \mathbf{F} \quad (\text{C.8})$$

So, if the matrix \mathcal{A} is invertible, the vector \mathbf{U} is equal to $\mathbf{U} = \mathcal{A}^{-1}\mathbf{F}$ and it is unique, which proves that the problem (C.4) has a unique solution. Matrix \mathcal{A} is called the *stiffness matrix*, referring to the mechanical problems when it was first introduced.

C.3 POISSON PROBLEM

The problem of Poisson can be written in the form:

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \quad (\text{C.9})$$

The 2D problem is being considered, i. e. Ω is an open bounded of \mathbb{R}^2 . Let us assume that the variational formulation of this problem admits a solution u in the space

$$V = \{u \text{ continuous and } C_1 \text{ piecewise on } \overline{\Omega}, u = 0 \text{ on } \partial\Omega\}$$

and try to approach it by a function u^h solution of the same variational problem, but where V is replaced by an approximation space V^h .

To describe the points of the domain, we need the two space variables x and y .

c.3.1 Choice of the mesh, the nodes and the local fields

The domain Ω is arbitrarily divided in triangles as shown on Fig C.1. The example is illustrated with 3 nodes per triangle and polynomials of degree 1 for the local field family. See Fig C.2 for the possibilities of elements in Actran. Each local field can then be determined with respect to the values at the 3 nodes. Note that using nodes at the summits of the triangles has two advantages:

1. The number of nodes is reduced, because there are common nodes to two adjacent elements
2. The C_0 continuity of the approximated solution is then ensured: the local fields of two adjacent elements will have the same value at the common node

Note that there is no obligation in providing each element with the same number of nodes: we could have taken elements with two nodes and others with three nodes for instance.

Fig C.1 shows that the mesh can be generated from a series of mappings (F_k) from a master triangle F_k to the subdomain Ω_k . Each local field can then be expressed in function of the values at the nodes. Indeed, there exists only one polynomial of degree 1, $v_k^h(x, y) =$

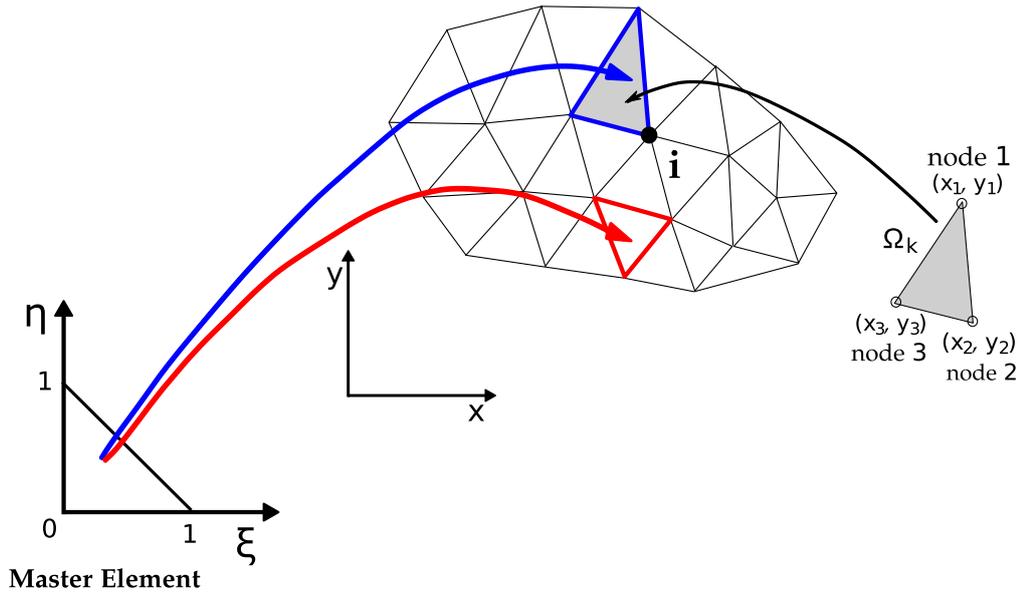


Figure C.1: A triangular finite element partition of a domain Ω into elements over which test functions are piecewise linear functions of the coordinates (x, y) .

$v^h|_{\Omega_k}(x, y) = ax + by + c$ which satisfies the following conditions:

$$\begin{cases} ax_1 + by_1 + c = v_1 \\ ax_2 + by_2 + c = v_2 \\ ax_3 + by_3 + c = v_3 \end{cases} \quad (\text{C.10})$$

where v_1, v_2 and v_3 denote the values of V^h at the nodes labelled 1, 2 and 3 respectively.

The system (C.10) can be rewritten:

$$\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (\text{C.11})$$

and is solved by inverting the 3x3 matrix:

$$\begin{aligned} \begin{pmatrix} a \\ b \\ c \end{pmatrix} &= \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\ &= \frac{1}{2A_k} \begin{pmatrix} y_2 - y_3 & y_3 - y_1 & y_1 - y_2 \\ x_3 - x_2 & x_1 - x_3 & x_2 - x_1 \\ x_2y_3 - x_3y_2 & x_3y_1 - x_1y_3 & x_1y_2 - x_2y_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \end{aligned}$$

where A_k represents the area of the triangle Ω_k , which is defined by

$$A_k = \begin{vmatrix} x_{k,1} & y_{k,1} & 1 \\ x_{k,2} & y_{k,2} & 1 \\ x_{k,3} & y_{k,3} & 1 \end{vmatrix} \quad (\text{C.12})$$

The local function can be written

$$v_k^h(x, y) = v_1 \psi_1^k(x, y) + v_2 \psi_2^k(x, y) + v_3 \psi_3^k(x, y) \quad (\text{C.13})$$

where the ψ_i are the element shape functions, which are given, according to the systems above:

$$\begin{cases} \psi_1^k(x, y) = \frac{1}{2A_k} [(x_2y_3 - x_3y_2) + (y_2 - y_3)x + (x_3 - x_2)y] \\ \psi_2^k(x, y) = \frac{1}{2A_k} [(x_3y_1 - x_1y_3) + (y_3 - y_1)x + (x_1 - x_3)y] \\ \psi_3^k(x, y) = \frac{1}{2A_k} [(x_1y_2 - x_2y_1) + (y_1 - y_2)x + (x_2 - x_1)y] \end{cases} \quad (\text{C.14})$$

If (x_i, y_i) are the coordinates of node i , and A_k is the area of Ω_k , the linear function has the property:

$$\psi_i^k(x_j, y_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (\text{C.15})$$

The elements are then all connected together to create the full computational domain Ω . By doing so, it creates the global piecewise

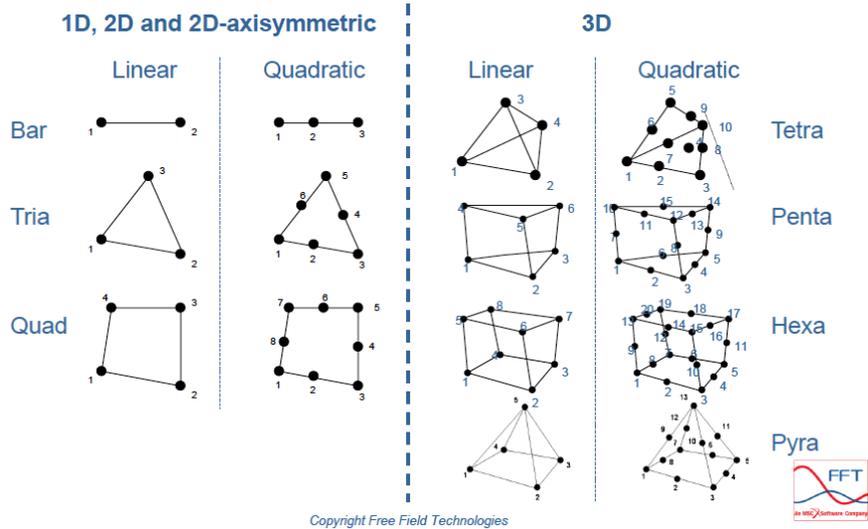


Figure C.2: Element topologies supported in Actran, from www.fft.be.

linear basis function ϕ_i , corresponding to each node i in the connected mesh. These *hat* functions are the result of grouping the element shape functions together, as shown in Fig C.3. The global test functions are linear combinations of these basis functions. For a mesh with M interior nodes,

$$v^h(x, y) = \sum_{i=1}^M v^h(x_i, y_j) \phi_i(x, y) \tag{C.16}$$

The process described with the *stiffness* matrix now needs to be performed, as in the subsection describing Garlekin Method.

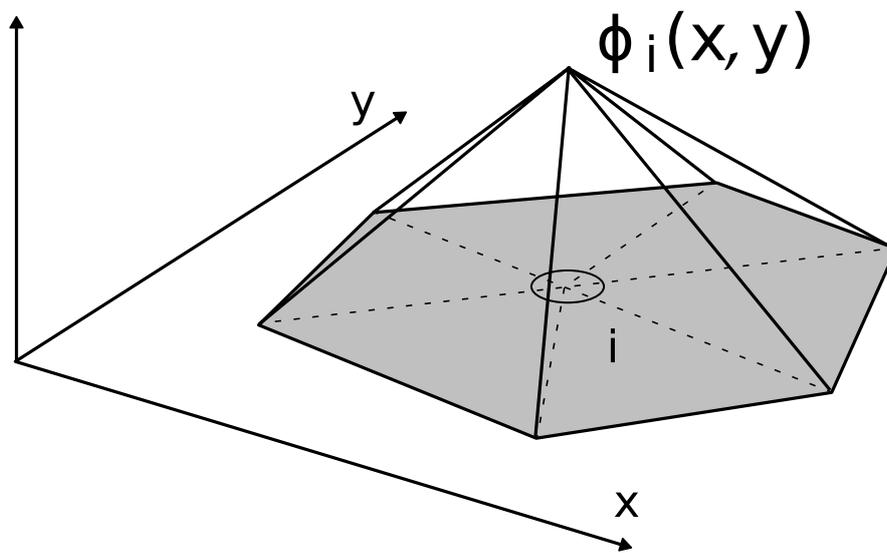


Figure C.3: A global basis function ϕ_i for the spaces U^h and V^h , created by assembling element shape functions which have the node i in common. The basis function is defined everywhere in Ω but is nonzero only on the subdomains surrounding the node i .

D | MRI

In 1974, Paul Lauterbur and Peter Mansfield independently published about *magnetic resonance imaging* (MRI) [43]. They were awarded with the Nobel Prize. Thirty years of development later, five MRI-related works have received the Nobel Prize.

The rise of the MRI would not have been possible without important findings in the fields of mathematics and physics, whose pioneers were [43]:

- **Jean Baptiste Joseph Fourier** (1768-1830) to whom we owe the Fourier transform, which enables a fast processing of the frequency and phase signals from the *nuclear magnetic resonance* (NMR) data yielding to the image reconstruction. In 1975, Richard Ernst was the first to use Fourier transform in the framework of MRI and it has been used ever since.
- **Nikola Tesla** (1856-1943) who discovered the rotating magnetic field which constitutes the basis of almost every alternative-current based machinery nowadays.
- **Sir Joseph Larmor** (1857-1942) from whom we inherited the Larmor equation which states that the frequency precession (i.e. spinning on its axis) of the nuclear magnetic moment ω is proportional to the product of the magnetic field amplitude B_0 and the gyromagnetic ratio γ : $\omega = \gamma B_0$. The Larmor equation is relevant in NMR because it gives the frequency at which the nucleus will absorb/emit energy, when immersed in a magnetic field. When the nucleus absorbs this energy, it will result in an alteration of its alignment.

Thanks to these important findings, the path to the MRI as we know it nowadays could be traced with important steps such as in [43]:

- 1937, **Isidor Rabi** observed and measured the magnetic moments of nuclei (1 in Fig D.1).
- 1946, **Edward M. Purcell** and **Felix Bloch** described independently of each other the phenomenon known as *magnetic resonance*: when certain nuclei are placed in a magnetic field, they absorb energy in the electromagnetic spectrum, which they re-emit once they return to their original state. (2 and 3 in Fig D.1).
- 1971, **Raymond Damadian** discovered that the nuclei of Hydrogen in cancerous tissues take a longer time to return to their original state compared to those from healthy tissues.

As a final step towards the beginning of the MRI Era, **Paul Lauterbur** and **Peter Mansfield**, without knowledge of each other's work, described how it was possible to locate NMR spatially with a magnetic field gradient (4 in Fig D.1)). The MRI was born.

D.1 BASIC PRINCIPLES

Magnetic Resonance Imaging, or **MRI** is a method used to get an image of the interior of a structure (tissue, bone, ...). An MRI Scan is composed of a magnet, magnetic gradient coils, a radio-frequency transmitter and receiver and a computer that control the RF (radio-frequency) emission and reception, acquisition of signals and imaging. If an atomic Nucleus is exposed to a static Magnetic field, it Resonates when a varying magnetic field is applied at the proper frequency [22]. This phenomenon is called the **Nuclear Magnetic Resonance** or **NMR**. Each nucleus in resonance sends a signal whose frequency and phase allow a reconstruction of the sources spatially and an image can be computed. This is the **MRI**.

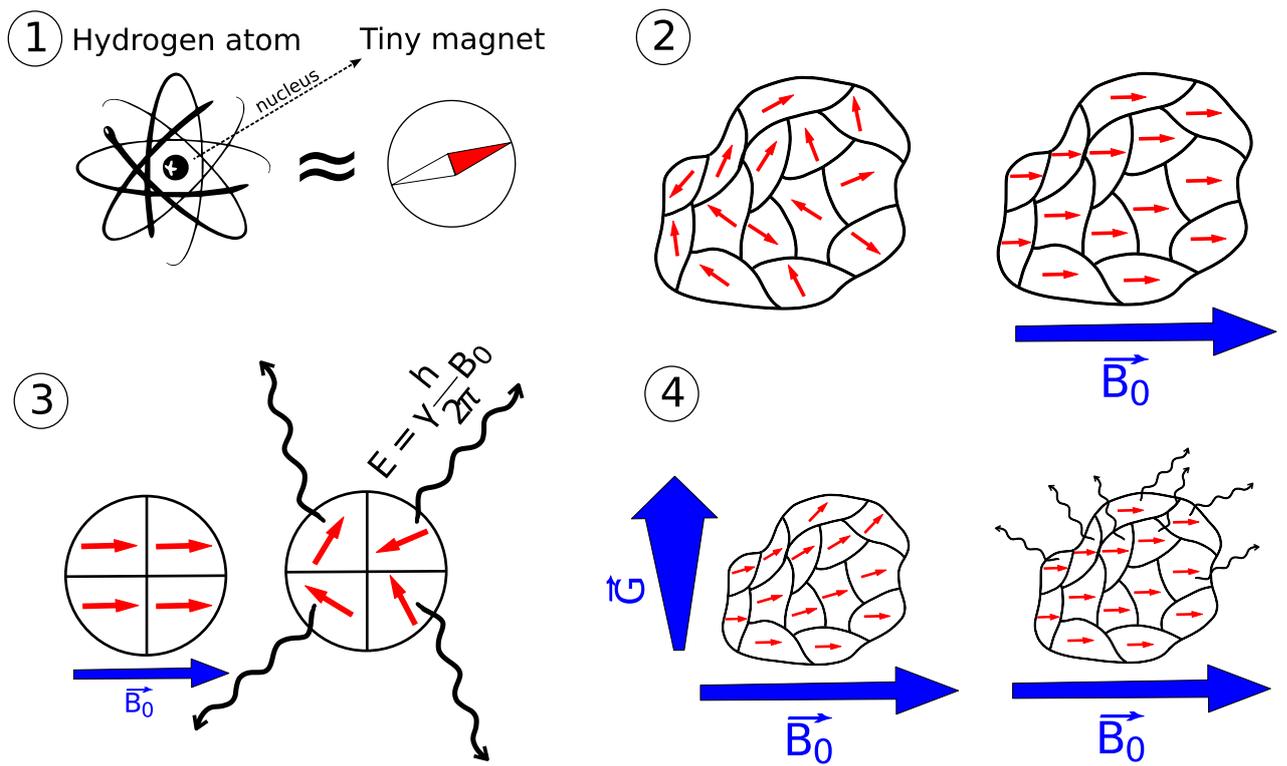


Figure D.1: NMR: processes at work.

In nature, the nuclei of many atoms, like Hydrogen, are behaving like tiny magnets (1 in Fig D.1). Bathed in the Earth's magnetic field, they tend to line up (2 in Fig D.1). Being composed of 2/3 of water, the human body contains a large number of atoms of Hydrogen. Therefore, the simple fact of walking on Earth makes the tiny compasses all over our body precessing (spinning on their axes like gyroscopes), aligned with the Earth's magnetic field (2 in Fig D.1). To create an MR image, this tendency of the nuclei to line up in the direction of an external magnetic field can be measured and changed [22]. By applying a gradient magnetic field, it is possible to act on the precession of the nuclei in the different part of the body, giving them a different frequency (their magneto-resonance frequency) according to their spatial location (4 in Fig D.1). After relaxing the nuclei, i.e. bringing them to their original state, they will release the energy absorbed at a certain frequency, yielding radio signals that are location dependent (3 in Fig D.1). From these signals, it is possible to compute and reconstruct a spatial image of the different tissues in the body.

When bathed in an external magnetic field B_0 , the atomic nucleus has an energy E which depends on its orientation. This energy is related to the resonance frequency ν by Planck's constant h :

$$E = h\nu \quad (\text{D.1})$$

and determines the strength of the signal which will be re-emitted once the nucleus returns to its steady state. The amount of energy E depends on the amplitude of the magnetic field B_0 and of a magnetic property of the nucleus called the gyromagnetic ratio γ . A fundamental equation of NMR and MRI links the resonance frequency ν of a nucleus to its gyromagnetic ratio γ when it is immersed in an external magnetic field B_0 . This is called *Larmor equation*:

$$\nu = \frac{\gamma}{2\pi} B_0 \quad (\text{D.2})$$

Combined with (D.1), this gives the energy absorbed/radiated by a nucleus of gyromagnetic ratio γ when it is plunged into an external magnetic field B_0 :

$$\boxed{E = h\nu = \gamma \frac{h}{2\pi} B_0} \quad (\text{D.3})$$

To the static field B_0 is added a transient field B_1 , which, when switched off, let the nuclei return to their steady state by emitting a radio wave whose energy is given by (D.3) when B_0 is replaced by B_1 . On the many atoms nuclei that are susceptible to NMR, just a few emit a signal strong enough to be processed in reconstructing an image in MRI. The nucleus of Hydrogen has the largest gyromagnetic ratio and, due to (D.3), has the highest energy and therefore the largest signal at any given field strength [22]. The human body is composed of 2/3 of water (H_2O). Hence there is a large abundance of Hydrogen nuclei in the human body. This, combined with the fact that it displays the strongest signal makes it the perfect candidate to get a high resolution.

To obtain the spatial location of the emission of the radio wave, it is necessary to superimpose a spatial gradient magnetic field $\mathbf{G} = (G_x, G_y, G_z)$ to the static field B_0 so that each nucleus will be immersed in a unique magnetic field, which gives the possibility of uniquely determining the spatial location of the nucleus according to the radio wave received in each of the three directions (x, y, z) ; (D.3) is modified accordingly:

$$\Delta E = h\nu = \gamma \frac{h}{2\pi} (B_0 + G) \quad (\text{D.4})$$

for each direction (x, y, z) . Since the atomic nuclei in different positions experience different values of $(B_0 + G)$, the image will be reconstructed by interpreting the spectral frequencies.

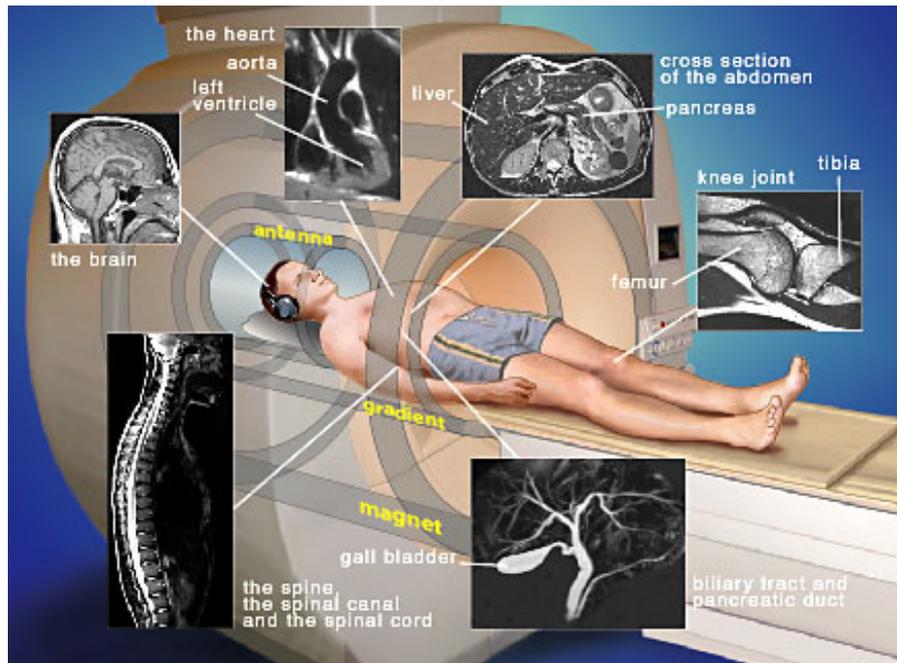


Figure D.2: http://www.nobelprize.org/nobel_prizes/medicine/laureates/2003/illpres/ Diagram of an MRI system.

D.2 ACQUISITION AND PROTOCOL

A major difference of MRI compared to all other human imaging techniques resides in the high quality of contrast especially amongst soft tissues [22]. The contrast of a typical MR image depends on the nuclei density and the exponential relaxation times of the signals consequent to the transient B_1 pulse. Different tissues may exhibit different proton concentration and/or different proton relaxation times. This enabled **Raymond Damadian** to differentiate cancerous tumours from healthy tissues [43]. The relaxation time T_1 (for spin lattice relaxation) is ~ 1 s for biological water protons: it is the time constant for the system to return to its thermal equilibrium [22], during which the the atomic nuclei are relaxing and loose the energy acquired by the transient B_1 to their environment (the lattice). However, the signals usually become too weak before the complete duration of T_1 because they dephase or become incoherent: this process is called spin-spin or T_2 relaxation and is ~ 1 s for biological water protons [22]. Therefore,

depending on the diagnoses, one can choose to weigh the parameters T_1 and T_2 to offer the best contrast for the image of the object under examination.

BIBLIOGRAPHY

- [1] Paavo Alku. Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering, 1992.
- [2] T. Arai. The Replication of Chiba and Kajiyama's Mechanical Models of the Human Vocal Cavity. *Journal of the Phonetic Society of Japan*, pages 31–38, 2001.
- [3] G. Ashalatha, P.R.; Deepa. *Textbook of Anatomy and Physiology for Nurses*. Jaypee Brothers Medical Publishers, 3rd edition, 2011.
- [4] B S Atal and Suzanne L Hanauer. Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. *Journal of the Acoustical Society of America*, 50:637–655, 1971.
- [5] T Baer, J C Gore, L C Gracco, and P W Nye. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America*, 90:799–828, 1991.
- [6] Wilmer T. Bartholomew. A Physical Definition of Good Voice-Quality in the Male Voice, 1934.
- [7] Alexander Graham Bell. *The Mechanism of Speech*. Funk & Wagnalls, 1910.
- [8] J. Berndtsson, G.; Sundberg. Perceptual significance of the center frequency of singer's formant. Technical report, KTH, 1994.
- [9] Erik Bresch and Shrikanth Narayanan. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. Technical report, 2010.

- [10] D Briche, G Malka, and J F Chassagne. The vallecule: an anatomic reality? *Revue de stomatologie et de chirurgie maxillo-faciale*, 92:213–218, 1991.
- [11] M Bunch and J Chapman. Taxonomy of singers used as subjects in scientific research. *Journal of voice : official journal of the Voice Foundation*, 14:363–369, 2000.
- [12] P. Castelli, E. and Badin. Vocal tract transfer functions with white noise excitation application to the naso-pharyngeal tract. In *Proc., 7th FASE Symposium, Edinburgh*, pages 415–422, 1988.
- [13] M Y Chen. Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, 102:2360–2370, 1997.
- [14] M. Chiba, T. and Kajiyama. *The Vowel-Its Nature and Structure*. Tokyo-Kaiseikan.
- [15] Philippe Clément, Stéphane Hans, Dana M Hartl, Shinji Maeda, Jacqueline Vaissière, and Daniel Brasnu. Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. A preliminary study., 2007.
- [16] Thomas F. Cleveland. Acoustic properties of voice timbre types and their influence on voice classification. *The Journal of the Acoustical Society of America*, 61(6):1622, June 1977.
- [17] J. P. Dalmont, C. J. Nederveen, and N. Joly. Radiation impedance of tubes with different flanges: numerical and experimental investigations. *Journal of Sound and Vibration*, 244(3):505–534, 2001.
- [18] J Dang and K Honda. Acoustic characteristics of the piriform fossa in models and humans. *The Journal of the Acoustical Society of America*, 101:456–465, 1997.

- [19] Y. Dang, J.; Shadle, C.H.; Kawanishi, Y.; Honda, K.; Suzuki. An experimental study of the open end correction coefficient for side branches within an acoustic tube. *J. Acoust. Soc. Am.*, 104(2):1075–1084, 1998.
- [20] Damian Murphy David Howard. *Voice Science, Acoustics, and Recording*. Plural Press, San Diego, 2008.
- [21] L Davies, R McGowan, and C Shadle. Practical flow duct acoustics applied to the vocal tract. In *Vocal Fold Physiology: Frontiers in Basic Science*, pages 93–142. 1993.
- [22] Joan Dawson and Paul Lauterbur. Magnetic resonance imaging. *Scholarpedia*, 3(7):3381, July 2008.
- [23] R. F. Detweiler. An investigation of the laryngeal system as the resonance source of the singer's formant. *Journal of Voice*, 8:303–313, 1994.
- [24] P. Djeradi, A.; Guerin, A.; Badin, B. and Perrier. Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method. *J. Phonetics*, 19:387–395, 1991.
- [25] A. Dmitriev, L. & Kiselev. Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities. *Folia Phoniatic*, 31:238–241, 1979.
- [26] Annette Dowd, John Smith, and J O E Wolfe. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time . Key words. *Language and Speech*, 41:1–20, 1998.
- [27] Matthias Echternach, Johan Sundberg, Tobias Baumann, Michael Markl, and Bernhard Richter. Vocal tract area functions and formant frequencies in opera tenors' modal and falsetto registers. *The Journal of the Acoustical Society of America*, 129:3955–3963, 2011.

- [28] J Epps, J R Smith, and J Wolfe. A novel instrument to measure acoustic resonances of the vocal tract during phonation. *Measurement Science and Technology*, 8(10):1112–1121, October 1997.
- [29] Robert Eymard, Herbin, and Thierry Gallouët. Finite volume method. *Scholarpedia*, 5(6):9835, 2010.
- [30] Raphaële Eymard, Robert and Gallouët, Thierry and Herbin. bookevol.pdf. In *Hanbook of Numerical Analysis*, pages 713–1020. 1997.
- [31] G Fant and M Bå vegård. Parametric model of VT area functions : vowels and consonants. *Speech, Music and Hearing-Quarterly Progress ...*, 1997.
- [32] Gunnar Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, second pri edition, 1960.
- [33] Gunnar Fant and S Pauli. Spatial characteristics of vocal tract resonance modes. In *Proceedings of the Speech Comm. Sem.*, pages 121–132, 1975.
- [34] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *PreprintsAudio Engineering Society*, 108:1–24, 2000.
- [35] Angelo Farina. Advancements in impulse response measurements by sine sweeps. In *New York*, volume 122, pages 4–9. Audio Engineering Society, 2007.
- [36] G Feng and E Castelli. Some acoustic features of nasal and nasalized vowels: a target for vowel nasalization. *The Journal of the Acoustical Society of America*, 99:3694–3706, 1996.
- [37] B. R. Fink. Folding Mechanism of the Human Larynx. July 1974.
- [38] J. L. Flanagan. *Speech analysis: Synthesis and Perception*. Springer-Verlag, 2nd editio edition, 1972.

- [39] N. H. Fletcher. Autonomous vibration of simple pressure-controlled valves in gas flows, 1993.
- [40] O Fujimura and J Lindqvist. Sweep-tone measurements of vocal-tract characteristics. *The Journal of the Acoustical Society of America*, 49:Suppl 2:541+, 1971.
- [41] Maëva Garnier, Nathalie Henrich, John Smith, and Joe Wolfe. Vocal tract adjustments in the high soprano range. *The Journal of the Acoustical Society of America*, 127(6):3771–80, June 2010.
- [42] J. Gaurrigues. La méthode des éléments finis. Technical report, 2002.
- [43] Tal Geva. Magnetic Resonance Imaging: Historical Perspective. *Journal of Cardiovascular Magnetic Resonance*, 8:573–580, 2006.
- [44] Stan Guy-bart, Embrechts Jean-jacques, and Archambeau Dominique. Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50:1–17, 2002.
- [45] Nathalie Henrich, John Smith, and Joe Wolfe. Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *The Journal of the Acoustical Society of America*, 129(2):1024–1035, February 2011.
- [46] A D Hillel. The study of laryngeal muscle activity in normal human subjects and in patients with laryngeal dystonia using multiple fine-wire electromyography. *The Laryngoscope*, 111:1–47, 2001.
- [47] J Hillenbrand, L A Getty, M J Clark, and K Wheeler. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97:3099–3111, 1995.
- [48] Kiyoshi Honda, Tatsuya Kitamura, Hironori Takemoto, Seiji Adachi, Parham Mokhtari, Sayoko Takano, Yukiko Nota, Hiroyuki Hirata, Ichiro Fujimoto, Yasuhiro Shimada, Shinobu

- Masaki, Satoru Fujita, and Jianwu Dang. Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling. *Computer methods in biomechanics and biomedical engineering*, 13:443–453, 2010.
- [49] Masaaki Honda. Human Speech Production Mechanisms. *Ntt Technical Review*, 1:24–29, 2003.
- [50] Hiroshi Imagawa, Ken-Ichi Sakakibara, Niro Tayama, and Seiji Niimi. The effect of the hypopharyngeal and supra-glottic shapes on the singing voice. In *Proceedings of the Stockholm Music Acoustics Conference*, volume 2003, pages 3–6, 2003.
- [51] Uno Ingard. *On the Design of Acoustic Resonators*, 1953.
- [52] Taisuke Itoh, Kazuya Takeda, and Fumitada Itakura. Acoustic analysis and recognition of whispered speech. *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1:I–389–I–392, 2002.
- [53] Elodie Joliveau, John Smith, and Joe Wolfe. Acoustics: tuning of vocal tract resonance by sopranos. *Nature*, 427:116, 2004.
- [54] K J Kallail and F W Emanuel. Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *Journal of speech and hearing research*, 27:245–251, 1984.
- [55] C. Kaltz, B.; D’Alessandro. Measurement of 3D phoneme-specific radiation patterns in speech and singing, LIMSI.
- [56] Ibrahim Kamal. Acoustic pharyngometry patterns of snoring and obstructive sleep apnea patients. *Otolaryngology-head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*, 130:58–66, 2004.
- [57] T Kitamura, K Honda, and H Takemoto. Individual variation of the hypopharyngeal cavities and its acoustic effects. *Acoust. Sci. and Tech*, 26(1), 2005.

- [58] Malte Kob, Nathalie Henrich, Hanspeter Herzel, David Howard, Isao Tokuda, and Joe Wolfe. *Analysing and Understanding the Singing Voice: Recent Progress and Open Questions*, 2011.
- [59] Anne Maria Laukkanen, Jaromír Horáček, Petr Krupa, and Jan G. Švec. The effect of phonation into a straw on the vocal tract adjustments and formant frequencies. A preliminary MRI study on a single subject completed with acoustic results. In *Biomedical Signal Processing and Control*, volume 7, pages 50–57, 2012.
- [60] Sang-Hyuk Lee, Hee-Jun Kwon, Hyun-Jin Choi, Nam-Hun Lee, Sung-Jin Lee, and Sung-Min Jin. The singer’s formant and speaker’s ring resonance: a long-term average spectrum analysis. *Clinical and experimental otorhinolaryngology*, 1:92–96, 2008.
- [61] Harold Levine and Julian Schwinger. *On the Radiation of Sound from an Unflanged Circular Pipe*, 1948.
- [62] Don Lewis. Vocal Resonance. *The Journal of the Acoustical Society of America*, 8(2):91, 1936.
- [63] John Makhoul. Linear Prediction : A Tutorial Review. *Proceedings of the IEEE*, 63:561–580, 1975.
- [64] Q Meng, D Sen, S Wang, and L Hayes. Impulse response measurement with sine sweeps and amplitude modulation schemes, 2008.
- [65] D G Miller, A M Sulter, H K Schutte, and R F Wolf. Comparison of vocal tract formants in singing and nonperiodic phonation. *Journal of voice : official journal of the Voice Foundation*, 11:1–11, 1997.
- [66] Scott Moisik. *The Epilarynx in Speech*. PhD thesis, University of Victoria, 2013.

- [67] Parham Mokhtari, Tatsuya Kitamura, Hironori Takemoto, and Kiyoshi Honda. Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients. *Journal of Phonetics*, 35:20–39, 2007.
- [68] Parham Mokhtari, Hironori Takemoto, and Tatsuya Kitamura. Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches. *Speech Communication*, 50:179–190, 2008.
- [69] K Motoki, N Miki, and N Nagai. Measurement of sound-pressure distribution in replicas of the oral cavity. *The Journal of the Acoustical Society of America*, 92:2577–2585, 1992.
- [70] Swen Müller and Paulo Massarani. Transfer Function Measurement with Sweeps. *J.AES*, pages 443–471, 2001.
- [71] D. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley. Acoustic Modeling Using the Digital Waveguide Mesh. *IEEE Signal Processing Magazine*, 24, 2007.
- [72] T M Nearey. Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85:2088–2113, 1989.
- [73] S. Nomura, Y.; Yamamura, I.; Inawashiro. On the acoustic radiation from a flanged circular pipe. *Journal of the Physical Society of Japan*, 15:510–517, 1960.
- [74] Alan O Cinneide. *Phase-Distortion-Robust Voice-Source Analysis*. PhD thesis, 2012.
- [75] J. Oden. Finite element method. *Scholarpedia*, 5(5):9836, May 2010.
- [76] T Oshima and M Imano. A Full Finite-Volume Time-Domain Approach towards General-purpose Code Development for Sound Propagation Prediction with Unstructured Mesh. In *Inter-Noise*, 2008.

- [77] C Painter. The laryngeal vestibule and voice quality. *Archives of oto-rhino-laryngology*, 243:329–337, 1986.
- [78] Gordon E. Peterson and Harold L. Barney. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24:175–184, 1952.
- [79] Y. Pham Thi Ngoc. *Caracterisation acoustique du conduit vocal: fonctions de transfert acoustiques et sources de bruit*. PhD thesis, Institut National Polytechnique de Grenoble, 1995.
- [80] Friederike Roers, Dirk Mürbe, and Johan Sundberg. Voice classification and vocal tract of singers: a study of x-ray images and morphology. *The Journal of the Acoustical Society of America*, 125(1):503–12, January 2009.
- [81] Thomas Rossing. *Springer Handbook of Acoustics*. Springer, 2007.
- [82] Sandra M Rua Ventura, Diamantino Rui S Freitas, Isabel Maria a P Ramos, and João Manuel R S Tavares. Morphologic differences in the vocal tract resonance cavities of voice professionals: an MRI-based study. *Journal of voice : official journal of the Voice Foundation*, 27:132–40, 2013.
- [83] K. Sell. *The disciplines of vocal pedagogy: towards an holistic approach*. Ashgate, 2005.
- [84] R.T. Smith, B. & Sataloff. *Choral pedagogy*. Singular, 1999.
- [85] M. Speed. *Voice Synthesis Using the Three-Dimensional Digital Waveguide Mesh*. PhD thesis, University of York, 2012.
- [86] M Speed, D T Murphy, and D Howard. Modeling the Vocal Tract Transfer Function using a 3D Digital Waveguide Mesh. *IEEE Transactions on Audio, Speech, and Language Processing*, 22(2):453–464, 2014.
- [87] Susan Standring, Neil R Borley, Patricia Collins, Alan R Crossman, Jeremiah C Gatzoulis, Michael A Healy, David Johnson,

- Vishy Mahadevan, Richard LM Newell, and Caroline B Wigley. *Grays Anatomy: The Anatomical Basis of Clinical Practice*. Elsevier, 40th editi edition, 2009.
- [88] R. Stein, E.; Shakarchi. *Fourier Analysis: An introduction*. Princeton University Press, 2003.
- [89] K. Stevens. *Acoustic Phonetics*. The MIT Press, 1998.
- [90] B H Story, A M Laukkanen, and I R Titze. Acoustic impedance of an artificially lengthened and constricted vocal tract. *Journal of voice : official journal of the Voice Foundation*, 14:455–469, 2000.
- [91] B H Story, I R Titze, and E A Hoffman. Vocal tract area functions from magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 100:537–554, 1996.
- [92] B H Story, I R Titze, and E A Hoffman. Vocal tract area functions for an adult female speaker based on volumetric imaging. *The Journal of the Acoustical Society of America*, 104:471–487, 1998.
- [93] B H Story, I R Titze, and E A Hoffman. The relationship of vocal tract shape to three voice qualities. *The Journal of the Acoustical Society of America*, 109:1651–1667, 2001.
- [94] Brad H. Story. Using imaging and modeling techniques to understand the relation between vocal tract shape to acoustic characteristics. In *Proceedings of the Stockholm Music Acoustics Conference*, pages 435–438, 2003.
- [95] Brad H. Story. Vowel Acoustics for Speaking and Singing. *Acta Acustica united with Acustica*, 90(4):629–640, 2004.
- [96] Brad H Story. Synergistic modes of vocal tract articulation for American English vowels. *The Journal of the Acoustical Society of America*, 118:3834–3859, 2005.

- [97] Brad H Story. Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions. *The Journal of the Acoustical Society of America*, 119:715–718, 2006.
- [98] Brad H Story. Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002., 2008.
- [99] Brad H. Story. The Vocal Tract in Singing. In *Handbook of Singing*. Oxford University Press, 2014.
- [100] A M Sulter, D G Miller, R F Wolf, H K Schutte, H P Wit, and E L Mooyaart. On the relation between the dimensions and resonance characteristics of the vocal tract: a study with MRI. *Magnetic resonance imaging*, 10:365–373, 1992.
- [101] J Sundberg. Articulatory interpretation of the "singing formant". *The Journal of the Acoustical Society of America*, 55:838–844, 1974.
- [102] J Sundberg. *The science of the singing voice*. Northern Illinois University Press, Dekalb, Illinois, 1987.
- [103] Johan Sundberg. The singers formant revisited. *STL-QPSR*, 36:83–96, 1995.
- [104] Hironori Takemoto, Seiji Adachi, Tatsuya Kitamura, Parham Mokhtari, and Kiyoshi Honda. Acoustic roles of the laryngeal cavity in vocal tract resonance. *The Journal of the Acoustical Society of America*, 120:2228–2238, 2006.
- [105] Hironori Takemoto, Kiyoshi Honda, Shinobu Masaki, Yasuhiro Shimada, and Ichiro Fujimoto. Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *The Journal of the Acoustical Society of America*, 119:1037–1049, 2006.
- [106] Hironori Takemoto, Parham Mokhtari, and Tatsuya Kitamura. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. *The Journal of the Acoustical Society of America*, 128:3724–3738, 2010.

- [107] M. Takuya, O. ; Imano. A Full Finite-Volume Time-Domain Approach towards General-purpose Code Development for Sound Propagation Prediction with Unstructured Mesh. In *Internoise 2008*, 2008.
- [108] I R Titze and B H Story. Acoustic interactions of the voice source with the lower vocal tract. *The Journal of the Acoustical Society of America*, 101:2234–2243, 1997.
- [109] Ingo R. Titze. *Principles of voice production*. Prentice Hall, 1994.
- [110] Ingo R. Titze. Acoustic interpretation of resonant voice. *Journal of Voice*, 15:519–528, 2001.
- [111] Ingo R. Titze. Is there evidence of a second singer’s formant? *J Singing*, 59(4):329–331, 2003.
- [112] Ingo R. Titze. A theoretical study of Fo-F1 interaction with application to resonant speaking and singing voice. *Journal of Voice*, 18:292–298, 2004.
- [113] Manjul Tiwari. Speech acoustics: How much science? *Journal of natural science, biology, and medicine*, 3(1):24–31, January 2012.
- [114] K Tom, I R Titze, E A Hoffman, and B H Story. Three-dimensional vocal tract imaging and formant structure: varying vocal register, pitch, and loudness. *The Journal of the Acoustical Society of America*, 109:742–747, 2001.
- [115] Tomás Vampola, Anne-Maria Laukkanen, Jaromír Horáček, and Jan G Svec. 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*, 129:310–315, 2011.
- [116] Shiqian Wang. Singer’s high formant associated with different larynx position in styles of singing. *Journal of the Acoustical Society of Japan*, 7(6):303–314, 1986.

- [117] R Weiss, W S Brown, and J Morris. Singer's formant in sopranos: fact or fiction? *Journal of voice : official journal of the Voice Foundation*, 15:457–468, 2001.
- [118] Jared J. Wolf. Efficient Acoustic Parameters for Speaker Recognition, 1972.
- [119] Joe Wolfe, Maëva Garnier, and John Smith. Vocal tract resonances in speech, singing, and playing musical instruments. *HFSP journal*, 3(1):6–23, January 2009.
- [120] N. Yan and N. Vocal tract dimensional characteristics of professional female singers with different types of singing voices. In *4th International Conference on Biomedical Engineering and Information*, pages 1549–1552, 2011.
- [121] H. Yokota, T.; Sakamoto, S.; Tachibana. Visualization of sound propagation and scattering in rooms. *Acoust. Sci. and Tech*, 23:40–46, 2002.
- [122] W. R. Zemlin. W. R. Zemlin Memorial Website, 2010.