Theoretical and experimental study of vocal tract acoustics

Rémi Blandin

Communauté
UNIVERSITÉ Grenoble Alpes

Thèse

pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Signal Image Parole Télécoms**

Arrêté ministériel : 7 août 2006

Présentée par
Rémi BLANDIN

Thèse dirigée par **Xavier PELORSON** et
codirigée par **Annemie VAN HIRTUM**

préparée au sein du
laboratoire GIPSA-lab
dans **l'école doctorale Électronique, Électrotechnique, Automatique, Traitement du Signal (EEATS)**

Theoretical and experimental study of vocal tract acoustics

Thèse soutenue publiquement le **27 octobre 2016**,
devant le jury composé de:

Murray Campbell
University of Edinburgh, Rapporteur
Joël Gilbert
LAUM, Rapporteur
Oriol Guasch i Fortuny
Ramon Llull University, Examinateur (Président du Jury)
Rafael Laboissière
LPNC, Invité
Xavier Pelorson
Gipsa-lab, Directeur de thèse
Annemie Van Hirtum
Gipsa-lab, Directrice de thèse

UNIVERSITÉ DE GRENOBLE ALPES

ÉCOLE DOCTORALE EEATS

Électronique, Électrotechnique, Automatique, Traitement du Signal

T H È S E

pour obtenir le titre de

docteur en sciences

de l'Université Grenoble Alpes

Mention : SIGNAL IMAGE PAROLE TÉLÉCOMS spécialité

Présentée et soutenue par

Rémi BLANDIN

Theoretical and experimental study of vocal tract acoustics

Thèse dirigée par Xavier PELORSON et Annemie VAN HIRTUM

préparée au laboratoire GIPSA-Lab

soutenue le 27 octobre 2016

Jury :

Rapporteurs :	Murray Campbell	-	University of Edinburgh
	Joël Gilbert	-	LAUM
Examinateur :	Oriol Guasch i Fortuny	-	Ramon Llull University (Président du Jury)
Invité :	Rafael Laboissière	-	LPNC
Directeurs :	Xavier Pelorson	-	Gipsa-lab
	Annemie Van Hirtum	-	Gipsa-lab

Acknowledgments

Je tiens à remercier tous ceux qui m'ont accompagné pendant ce travail. Tout d'abord mes encadrants Xavier et Annemie qui m'ont donné l'opportunité de mener à bien ce travail de recherche et m'ont guidé et conseillé tout au long de ce parcours. Ensuite Rafael qui a commencé le travail de modélisation acoustique du conduit vocal longtemps auparavant avec Xavier et qui m'a permis de continuer ce projet. Je le remercie en particulier pour les heures passées à m'aider à comprendre le programme et à essayer de comprendre les bugs. Je tiens aussi à remercier Marc et Oriol de l'université La Salle de Barcelone avec qui nous avons pu effectuer une excellente collaboration. Les nombreux résultats issus des simulations éléments finis réalisé par Marc présentés ci-après attestent de la qualité et du bénéfice de cet échange scientifique. Je tiens également à remercier Alexander et Stefan de la FAU à Erlangen qui m'ont permis de collaborer avec eux de façon enrichissante. Je souhaite également remercier tous les autres membres du projet EUNISON, Saeed qui a extrait les geometries de conduit vocal des données IRM, Sten et Oriol qui ont géré ce projet et tous les autres qui ont apporté leurs contributions. Une grande partie des résultats expérimentaux présentés ci-après n'aurait pu être obtenue sans le travail de Xavier Laval (X ou Javier pour les intimes) qui a monté de nombreux setups, créé de nombreux programmes d'acquisition Labview et a toujours été très disponible pour aider à résoudre les problèmes expérimentaux. Je tiens à remercier également tous les stagiaires avec qui j'ai pu travailler et qui ont apporté leurs contributions plus ou moins importantes à ce travail, Martin, Florian, Guillaume, Marguerite, Michael, Stephane, Romain, Benoît, Clément, Elisa et William. Je voudrais aussi remercier ceux qui ont fait que cette expérience à également été agréable et enrichissante sur le plan humain. Merci aux différentes personnes qui ont partagé mon bureau, Serap pour les nombreuses sorties ski et les cookies, Jessy qui m'a accompagné à travers bois et avec qui on a transporté des machines à laver et Andrew pour ses discussions intéressantes et le partage de ses grandes connaissances sur de nombreux sujets. Je tiens à remercier aussi les autres personnes qui ont fréquenté le DPC. Javier avec qui je suis allé aux quatre coins de l'Europe (Turquie, Italie, Allemagne, Croatie, Espagne) et dont la bonne humeur et l'optimisme sont toujours appréciés. Diandra pour son sens de la repartie et qui sait mettre la bonne ambiance dans la cafet (même si elle ne boit pas de thé). Angélique pour son sens de l'écoute, les randos, les sorties ski et les échanges de yaourts. Gustavo pour les soirées passées à "travailler" au Loco Mosquito, son habileté à prélever la nourriture directement dans la nature en Corse et son amour pour les "gatos". Monalisa pour sa merveilleuse cuisine et son enthousiasme pour les randos. Jean Francois pour sa bonne humeur, son sans de l'humour omniprésent et ses blagues racontées à l'envers. Antje pour les sorties vélos et les soirées gastronomiques dans son grand appart. Sanjeev et Bharat pour la bonne cuisine indienne, les heures passées à jouer au ping pong et les glissades dans la neige. Cecilia, Elsa, Laure, Sonia, Fanny, Martine, Marion, Sandra et tous les autres membres de l'administration avec qui il a toujours été agréable d'interagir. Tous les autres qui ont participé à la bonne ambiance au DPC, Yo, Louis, Nicolas, Paul, Marie Lou, Jonathan, Alexandre, Tiphaine, Thibault, Clémence, Sari, Eric, Melaine, Li, Guillaume, Maël, Adela, et tous ceux que j'oublie. Je tiens aussi à remercier ma famille et mes amis qui ont été avec moi tout au long de cette expérience et qui ont préparé un merveilleux pot de soutenance. Et enfin, merci à toi Elena de m'avoir accompagné à la fin de ma thèse.

Contents

List of Figures

List of Tables

Table of acronyms

3D	*3 Dimensional*
FD	*Finite Differences*
FEM	*Finite Element Method*
HOM	*Higher Order Modes*
MM	*Multimodal Method*
MRI	*Magnetic Resonance Image*
MSPLD	*Maximal Sound Pressure Level Difference*
PPTF	*Pressure/Pressure Transfer Function*
VPTF	*volume Velocity/Pressure Transfer Function*

Chapter 1

Introduction

This work has been done in the framework of the EUNISON European project, which aims to build a detailed and accurate physical model of the voice.

1.1 State of the art

1.1.1 Vocal tract acoustics

1.1.1.1 The vocal tract

The vocal tract corresponds to the cavity located between the vocal folds and the lips. Its length and its width are of the order of 17 cm and 5 cm respectively for an adult. It is an important part of the vocal apparatus which act as an acoustic resonator and allows the speaker to create sound sources. It acoustically interacts with the different sound sources which generates speech and singing sounds: the air jet pulsed by the vocal fold oscillation and the turbulent sound sources generated by the constrictions present in the vocal apparatus.

A speaker or a singer can vary the shape of his vocal tract by moving his jaw, tongue and lips in order to change the acoustic properties of the vocal tract. This allows one to produce different phonemes. Some of them are mainly characterized by the resonances, denoted as formants in speech study, of the vocal tract. This is the case of the vowels which are identified and classified using the two first formants of the vocal tract. The vowels [ɑ], [i] and [u] constitute the extremes in term of variation of the formants, [ɑ] having the highest first fomant, [i] the highest second formant and [u] the lowest second formant. Thus, the acoustic properties of the vocal tract appear as critical for the differentiation of some phonemes and for the quality and the timbre of the speech and singing production.

In the case of the voiced sounds (the sound for which the sound source is created by the oscillations of the vocal folds), it is particularly interesting to characterize the contribution of the vocal tract with a transfer function which relates the particle velocity generated by the sound source at the entrance of the vocal tract to the acoustic pressure radiated outside of the vocal tract.

1.1.1.2 Simple modeling of vocal tract acoustics: the plane wave model

In order to study the acoustic properties of the vocal tract, it is useful to perform acoustic modelling. In most of the physical models used in speech and singing study, the frequency range is limited from 0 kHz to about 5 kHz. This limited frequency range which does not cover the entire range of perception of the human ear is often used because it is sufficient to ensure speech recognition and it allows one to use very simple acoustic models. In particular, one can assume that plane waves are travelling inside the vocal tract, which means that the acoustic pressure varies only along the propagation direction and depends only on the area of the cross-sections of the vocal tract [1]–[4]. Thus, in this case a one-dimensional description with an area function of the vocal tract shape is sufficient. This allows one to perform simulations at a very low computational cost: it is possible to perform real time simulations [5]. However, given the limited frequency range used, the quality of the simulations and synthesis is not optimal.

1.1.1.3 Speech directivity

Another aspect of the transfer function between the sound source and the radiated sound is the directivity effect. It consists in the variation of the amplitude and the phase of the sound radiated by a speaker with the direction. The directivity is studied for various purposes, including microphone placement optimisation [6], telephony [7], vocal performance practice [8]–[10], experimental validation of acoustic theories [11], architectural acoustics [12], auralization and three-dimensional (3D) sound synthesis [13], [14]. The observation of the directivity on real speakers or singers has shown that the directivity patterns become more complex and pronounced at high frequency [6], [8]–[10]. At low frequency (from 0 kHz to about 5 kHz), this phenomenon is modeled by accounting for the diffraction of the acoustic waves by the head and by taking into account the reflections on the shoulder and the torso. The pressure field at the mouth exit is modeled as a pulsating sphere.

1.1.2 Available supporting data

The limits of the simple modeling of the vocal tract acoustics and the necessity to have more accurate modeling have been highlighted in the literature.

1.1.2.1 Limits of the plane wave model

Recent findings in speech perception emphasized the importance of frequencies beyond 5 kHz [15]. On the other hand, new technologies and applications (such as wideband telephony, augmentative hearing or 3D sound synthesis) use frequencies beyond 5 kHz in order to improve the sound quality and the communication comfort. At frequencies higher than 5 kHz, the wavelength of the acoustic waves can become of the same order of size as the transverse

dimension of the vocal tract. In this case, the plane wave assumption does not hold because variations of the acoustic field in the plane perpendicular to the propagation direction can be present. This has been observed both by simulations accounting for the 3D aspect of the pressure field [16]–[18] and by measurements of the pressure field performed in vocal tract replicas [19], [20]. As a consequence of these 3D variations of the acoustic field inside the replicas, departures from the plane wave theory have been observed from around 5 kHz [16]–[18], [21], [22]. Thus, the plane wave assumption appears as too limited to be used beyond about 5 kHz and other acoustical models accounting for the 3D aspects of the pressure field are needed for this frequency range.

1.1.2.2 High frequencies in speech directivity

In what concerns the directivity of the radiated sound, for frequencies higher than about 5 kHz, the wavelength of the acoustic waves can be of the same order of length as the mouth aperture. As a consequence, the diffraction of the acoustic waves present at the mouth exit becomes non negligible. On the other hand, variations of the acoustic field over the plane of the mouth aperture can occur. Thus, the pulsating sphere model is no more accurate to predict the directivity of the radiated sound, and the acoustic field inside the vocal tract is likely to influence the directivity of the radiated sound. As a consequence, the directivity is expected to vary for different phonemes. Unfortunately, these variations are masked in most of the directivity measurements performed on real speakers or singers, because the directivity patterns are obtained from averaged spoken or sung sentences. However, it is shown in a study of Monson [10] that there are substantial differences between the directivity patterns of different fricatives. On the other hand, it has been reported in the same study that there is a difference of directivity pattern in high frequency which is related to gender, which may be explained by a difference of vocal tract size and shape.

1.2 Formulation of the problem

There are very few studies reporting experimental measurements of the pressure field inside vocal tract replicas. The progress of medical imagery [23]–[26] and of 3D printing allows one to build more accurate replicas than the one used in the study of Motoki [19]. So more accurate observations and measurements can be performed. On the other hand, the effects of some geometrical features of the vocal tract, such as the eccentricity, the cross-sectional shape or the mouth aperture, at high frequency are not clearly understood. The geometrical details necessary to take into account in order to obtain a qualitatively realistic behavior of simplified vocal tract geometries have not been identified. The 3D acoustic simulation methods, such as finite elements (FEM), finite differences (FD) and multimodal method (MM) needs to be compared and validated with experimental measurements. The influence of the internal acoustic field and of its 3D aspect has not been investigated and is usually not taken into account in the modelling of the speech directivity. In addition, no study has brought any explanation of the physical origin of the directivity differences related to the phonemes. On

the other hand, the directivity measurements have been all performed in octave bands or in third of octave bands and with a minimal angular resolution of 15°. This may mask some phenomena which have important variations within small frequency intervals and small angular regions.

1.3 Aim and outline of the thesis

The objectives of the work presented in this document are:

- To investigate how the 3D aspects of the acoustic field inside the vocal tract can affect the transfer function of the vocal tract and the directivity of the radiated sound, focusing on the influence of geometrical features of the vocal tract shape including the eccentricity, the cross-sectional shape, the size of the mouth aperture and the lips.
- To measure accurately acoustic pressure fields, transfer functions and directivity patterns with a thinner frequency, angular resolution (5Hz and 3°) than the one already performed on real speakers and singers (third of octave bands and15°) and to validate experimentally 3D acoustic simulation methods with it.
- To investigate how simplified vocal tract geometries can reproduce qualitatively the properties of realistic ones.
- To compare the acoustic properties of the vocal tract shape corresponding to three vowels which correspond to the extremes possible values of the formants ([ɑ], [i] and [u]).

In order to account for the 3D aspects of the acoustic field, one can use any simulation method which allows to take it into account. However, one has chosen for this work to apply the multimodal method [27]–[30] to the case of the vocal tract. Indeed, besides having a lower computational cost than the other simulation methods, it has the advantage to provide all the information related to the propagation modes (eigen functions, cut-on[1] frequencies, modal amplitudes and projection matrices). This is a valuable aid to understand the relationship between the geometrical features of the vocal tract shapes and the properties of the acoustic field. However, the implementation proposed here for this method is limited to straight waveguides[2] constituted of a concatenation of sections with arbitrary cross-section. Thus, this implementation is particularly interesting for the simplified geometries based on area function as an example. On the other hand FEM simulations have been used for more complex geometries.

Several vocal tract geometries with different degree of complexity have been used. Realistic replica based on magnetic resonance images (MRI) [26] have been used to investigate the

[1]The term cut-on frequency refers to the fact that a propagation mode begins to propagate above its cut-on frequency, it is the same as the cutoff frequency which refers to the fact that a propagation mode stops propagating under its cutoff frequency.

[2]Bent waveguides can however be described with this method [31].

acoustic properties, the directivity of realistic geometries and the influence of the lips. However, this type of geometry is subject to inter- and intra-speaker variations and their shape is complex, which makes the understanding of the ongoing phenomena more difficult. That is why, simplified geometries based on two [1] and 44 [23] tubes concatenations have been extensively studied. These geometries are fully described by a given area function, which allows one to create several versions with different geometrical details, such as the eccentricity of the junctions and the cross-sectional shape, and to study their influence.

In order to perform measurements of the acoustic pressure, mechanical replicas of some of these geometries have been built either with plexiglass tubes or using a 3D printer. An experimental setup allowing to measure the acoustic pressure has been designed. This allowed one to acquire experimental pressure fields, transfer functions and directivity patterns which have been compared to simulations performed with MM and FEM.

This document is organized as follows:

- In chapter 2 the MM implemented for this work is detailed.
- In chapter 3 the vocal tract geometries, the experimental setup, the data analysis methods and FEM are presented.
- In chapter 4 the simulated and measured transfer functions and pressure fields are presented and discussed.
- In chapter 5 the simulated and measured directivity patterns are presented and discussed.

Chapter 2

The multimodal approach applied to vocal tract

Contents

In this chapter the multimodal theory used to compute the acoustic pressure and velocity in a variable cross-section waveguide is detailed. This theory has been developed by several authors including Roure [27], Kergomard [28], Pagneux [29] and Kemp [30]. It has already been applied to the vocal tract case with rectangular cross-sections by Motoki [32]. The aforementioned works are extended to consider straight vocal tract geometries with arbitrary cross-sections and eccentric junctions. In order to solve the Helmholtz equation

$$\Delta p + k^2 p = 0 \ , \tag{2.1}$$

where Δ is the Laplacian operator, k is the wavenumber and p is the pressure, the strategy is to divide the waveguide in a succession of constant cross-section waveguides in which a simple solution can be derived. By applying the continuity equations of the pressure and the velocity at each junction and taking into account the radiation impedance at the open end of the waveguide, one can backpropagate the impedance or admittance matrices from the open end to the entrance. The acoustic pressure and velocity can then be propagated from the entrance toward the exit and computed in each section. In what follows the assumptions of linear acoustics (small perturbations assumption) are considered.

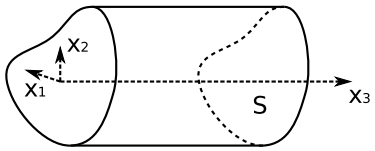


Figure 2.1: Constant cross-section waveguide.

2.1 Constant cross-section waveguide

In this section the multimodal theory is detailed in the case of a constant cross-section waveguide (see Fig. 2.1). Throughout this document, a generalized coordinate system (x_1, x_2, x_3) is used, x_3 being the propagation direction. The pressure p and the volume velocity u_{x_3} in the propagation direction x_3 inside a waveguide with arbitrary constant cross-sectional shape can be expressed as the summation of the contribution of an infinity of propagation modes ψ_n [33] weighted by a propagation factor $\left(A_n e^{-jk_n x_3} \pm B_n e^{jk_n x_3}\right)$ and a time evolution factor $e^{j\omega t}$,

$$\begin{cases} p(x_1,x_2,x_3,t) &= \sum\limits_{n=0}^{\infty} \psi_n(x_1,x_2)\left(A_n e^{-jk_n x_3} + B_n e^{jk_n x_3}\right) e^{j\omega t} \ , \\ u_{x_3}(x_1,x_2,x_3,t) &= \frac{S}{\rho c}\sum\limits_{n=0}^{\infty} \psi_n(x_1,x_2)\left(A_n e^{-jk_n x_3} - B_n e^{jk_n x_3}\right) e^{j\omega t} \ , \end{cases} \tag{2.2}$$

where S is the surface of the cross-section, ρ is the constant air density, c is the speed of sound at rest, A_n and B_n are the amplitude of the waves propagating forward and backward respectively and ω is the angular frequency. To each propagation mode ψ_n corresponds a modal wavenumber k_n in the propagation direction x_3 which satisfies the dispersion relationship

$$k^2 = k_{x_1}^2 + k_{x_2}^2 + k_n^2 \ , \tag{2.3}$$

where $k = \omega/c$ is the free field wavenumber. When no particular cross-sectional shape is implied a single subscript n is used to identify the different propagation modes ψ_n. In the particular cases described in sections 2.1.1.1, 2.1.1.2 and 2.1.1.3, two or three subscripts are used because they are introduced in their expressions and it is meaningful in what concerns the mode-shapes.

To avoid heavy expressions it is more convenient to use a vector notation (throughout this document vectors and matrix are in bold) and to omit the $e^{j\omega t}$ factor:

$$\begin{cases} p(x_1,x_2,x_3) &= \boldsymbol{\psi^t P} \\ u_{x_3}(x_1,x_2,x_3) &= \boldsymbol{\psi^t U} \ , \end{cases} \tag{2.4}$$

where the superscript t represent the transpose operator,
$\boldsymbol{\psi} = [\psi_1(x_1,x_2), \psi_2(x_1,x_2), ..., \psi_n(x_1,x_2)]^t$ is a vector containing the magnitude of the propagation modes ψ_n at the location (x_1,x_2,x_3) and the time instant t, $\boldsymbol{P} = [\, A_1 e^{-jk_1 x_3} + B_1 e^{jk_1 x_3}, A_2 e^{-jk_2 x_3} + B_2 e^{jk_2 x_3}, ..., A_n e^{-jk_n x_3} + B_n e^{jk_n x_3} \,]^t$ is a vector containing the amplitude of the propagation modes ψ_n for the acoutic pressure, $\boldsymbol{U} = \frac{S}{\rho c}[\, \left(A_1 e^{-jk_1 x_3} - B_1 e^{jk_1 x_3}\right), \left(A_2 e^{-jk_2 x_3} - B_2 e^{jk_2 x_3}\right), ..., \left(A_n e^{-jk_n x_3} - B_n e^{jk_n x_3}\right) \,]^t$ is a vector containing the amplitude of the propagation mode ψ_n for the volume velocity.

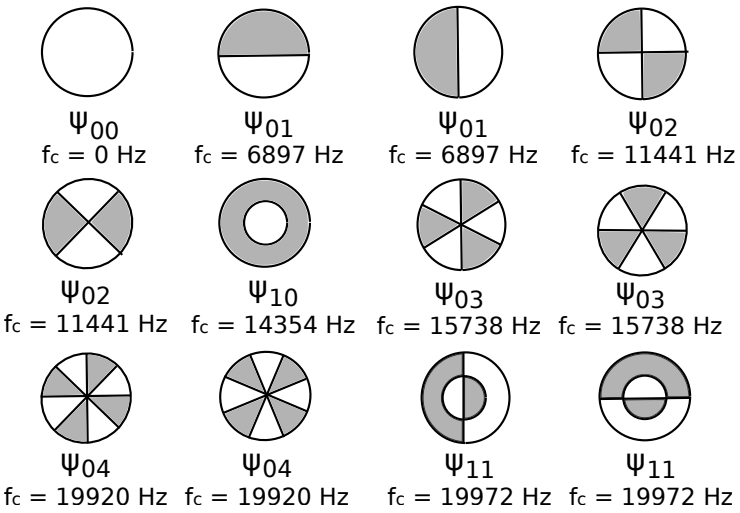


Figure 2.2: Mode-shapes and cut-on frequencies f_c (computed with Eq. (2.8)) of the propagation modes of a circular cross-section whose cut-on frequency is below 20 kHz for a diameter of 29.5 mm and a temperature of 26.5°C. The lines are zero amplitude nodal lines and the color indicates the phase change.

2.1.1 Propagation modes

The functions ψ_n are the solutions of the two dimensional Helmholtz equation. They are part of an orthogonal modal basis and they satisfy a normalisation condition

$$\int_S \psi_m^* \psi_n dS = S\delta_{mn} \ , \tag{2.5}$$

where S is the cross-section surface.

Each propagation mode can propagate above its cut-on frequency f_c. The mode of order $n = 0$, ψ_0, is the plane mode, the only one typically considered in vocal tract acoustics. Its cut-on frequency being 0 Hz, it can propagate at any frequency. The other modes are denoted as Higher Order Modes (HOM) in this document. Their cut-on frequency is higher than 0 Hz, and below this frequency they are exponentially damped along the propagation axis from the excitation point. In this case they are denoted as evanescent.

For the practical implementation of the MM approach, the infinite summation of Eq. (2.2) needs to be truncated to $N + 1$ modes. In the implementation used for this work, it has been chosen to use only the modes whose cut-on frequency lies under an arbitrary frequency limit (typically 20 kHz). As a consequence, the number of HOM used depends on the cross-sectional shape. Thus, in the case of a complex geometry with multiple sections, it varies according to the considered section, more propagation modes being considered in a section having a large cross-section than in one with a small cross-section.

2.1.1.1 Circular cross-sections

In the case of a circular cross-section of radius R, the propagation modes can be expressed using polar coordinates (r, θ) in the plane (x_1, x_2) as a product of Bessel, sine and cosine functions

$$\psi_{mn}(r,\theta) = \frac{J_n(r\gamma_{mn}/R)}{N_{mn}} \begin{cases} \sin(n\theta) \\ \cos(n\theta) \end{cases}, \tag{2.6}$$

where

$$N_{mn} = \begin{cases} \sqrt{J_n(\gamma_{mn})^2 - J_{n-1}(\gamma_{mn})J_{n+1}(\gamma_{mn})} & : \quad n = 0, \\ \frac{1}{2}\sqrt{J_n(\gamma_{mn})^2 - J_{n-1}(\gamma_{mn})J_{n+1}(\gamma_{mn})} & : \quad n > 0, \end{cases} \tag{2.7}$$

is a normalization constant (for the computation of N_{mn} see appendix B), γ_{mn} is the order m zero of the first derivative of the Bessel function J_n. The cut-on frequency f_c of each mode ψ_{mn} is given by

$$f_{cmn} = \frac{c\gamma_{mn}}{2\pi R}. \tag{2.8}$$

The mode-shapes and the cut-on frequencies of a circular cross-section corresponding to some geometries studied later are presented for $f_c < 20$ kHz in Fig. 2.2. They correspond to a diameter of 29.5 mm and the cut-on frequencies have been computed for a temperature of 26.5°C which corresponds to an experiment whose results are presented later (see section 4.1). The modes ψ_{01}, ψ_{02}, ψ_{03}, ψ_{04} and ψ_{11} are degenerated: they have two mode-shapes similar by rotation of a certain angle having the same cut-on frequency. One can see that the first subscript index corresponds to the radial dimension and the second to the angular dimension.

2.1.1.2 Rectangular cross-sections

In the case of a rectangular cross-section of dimensions a in x_1 and b in x_2, the functions ψ_{mn} can be expressed as

$$\psi_{mn}(x_1,x_2) = M_m \cos\left(\frac{m\pi x_1}{a}\right) N_n \cos\left(\frac{n\pi x_2}{b}\right) \tag{2.9}$$

with $\begin{cases} M_m = 1: & m = 0, \\ M_m = \sqrt{2}: & m > 0, \end{cases} \quad \begin{cases} N_n = 1: & n = 0, \\ N_n = \sqrt{2}: & n > 0. \end{cases}$

Their cut-on frequencies are given by

$$f_{cmn} = \sqrt{\left(\frac{mc}{2a}\right)^2 + \left(\frac{nc}{2b}\right)^2}. \tag{2.10}$$

The mode-shapes and the cut-on frequencies of a square cross-section corresponding to some geometries studied later are presented for $f_c < 20$ kHz in Fig. 2.3. They correspond to a width of 29.5 mm and the cut-on frequencies have been computed for a temperature of 26.5°C. The modes ψ_{01}, ψ_{10}, ψ_{02}, ψ_{20}, ψ_{12}, ψ_{21}, ψ_{03}, ψ_{30}, ψ_{13} and ψ_{31} are degenerated: their mode-shapes are similar by a rotation of 90°and they have the same cut-on frequency. One can see that the first subscript index corresponds to the number of nodal lines along the axis x_1 and the second to the number of nodal lines along the axis x_2.

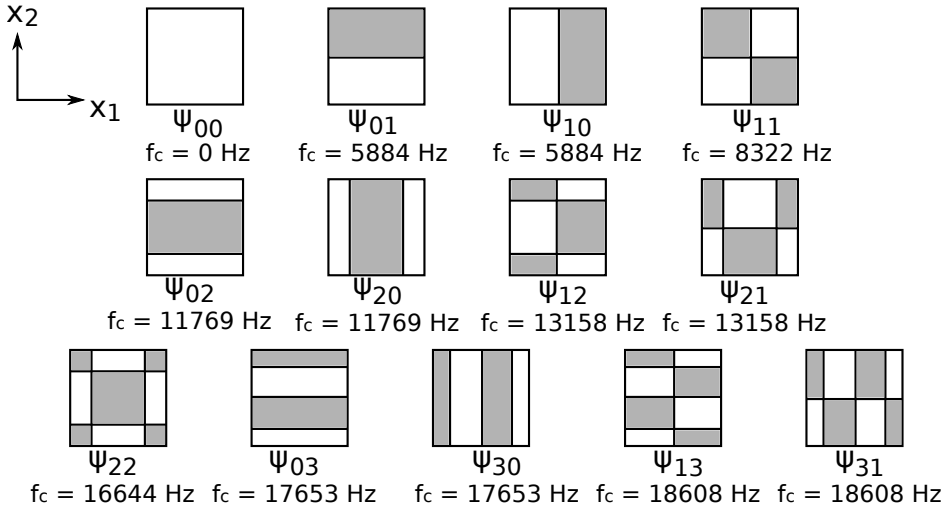


Figure 2.3: Mode-shapes and cut-on frequencies f_c (computed with Eq. (2.10)) of the propagation modes of a square cross-section whose cut-on frequency is below 20 kHz for a width of 29.5 mm and a temperature of 26.5°C. The lines are zero amplitude nodal lines and the color indicates the phase change.

2.1.1.3 Elliptical cross-sections

The case of an elliptical cross-section can be solved analytically using elliptical coordinates (ξ,η). For a full devellopement of this solution see [34] and [35]. The elliptical coordinates can be related to the coordinates (x_1,x_2) with the relations

$$x_1 = h\cosh\xi\cos\eta \ , \tag{2.11}$$

$$x_2 = h\sinh\xi\sin\eta \ , \tag{2.12}$$

where h is the semi inter-focal distance of the ellipse. The boundary of the elliptical cross-section is defined by $\xi = \xi_{max}$, with $x_{1max} = 2h\cosh(\xi_{max})$ and $x_{2max} = 2h\sinh(\xi_{max})$ the half lengths of the major and minor axis respectively. The resolution of the Helmholtz equation (Eq. (2.1)) on the cross-section leads to the Mathieu equation

$$\frac{d^2V}{d\eta^2} + (\sigma - K^2\cos^2\eta)V = 0 \ , \tag{2.13}$$

and the modified Mathieu equation

$$\frac{d^2U}{d\xi^2} - (\sigma - K^2\cosh^2\xi)U = 0 \ , \tag{2.14}$$

with $\psi(\xi,\eta) = U(\xi)V(\eta)$, σ the eigenvalue of the Mathieu function and $K^2 = h^2[(\omega/c)^2 - k_n^2]$, where k_n is the wavenumber in the propagation direction x_3. Two kinds of functions, denoted as even and odd, are derived from Eqs. (2.13) and (2.14) [36], [37]:

$$\psi_{emn}(\xi,\eta) = \frac{1}{N_{emn}} S_{em}(K_{emn},\cos(\eta)) J_{em}(K_{emn},\cosh(\xi)) \ , \tag{2.15}$$

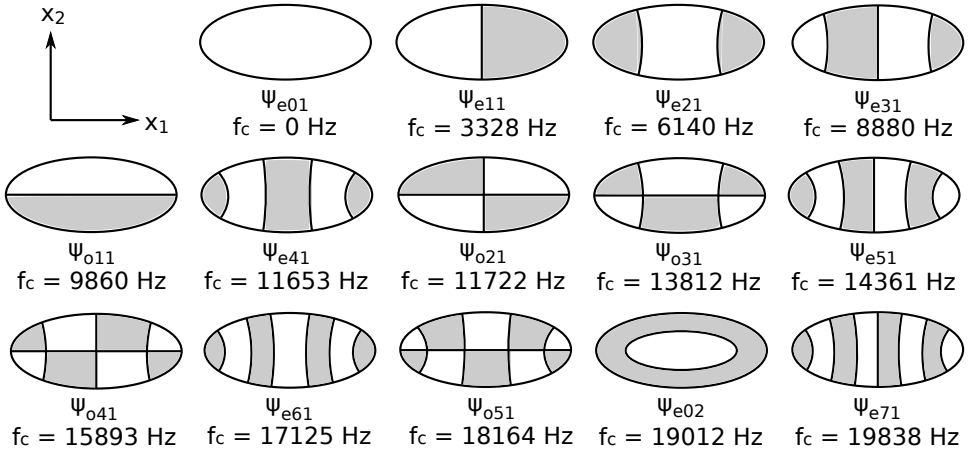


Figure 2.4: Propagation modes whose cut-on frequency f_c lies under 20 kHz for the dimensions corresponding to the widest part of a vocal tract geometry with elliptical cross-sections (6.2 cm wide and 1.9 cm thick) and a temperature of 22.65 °C. The lines are zero amplitude nodal lines and the color indicates the phase change.

where the ubscript e stands for even, N_{emn} is a normalization constant computed so that Eq. (2.5) is satisfied, S_{em} is the even Mathieu function of the first kind of order m [38], J_{em} is the even modified Mathieu function of the first kind of order m, $m \in \mathbb{N}, m \geq 0, n \in \mathbb{N}, n > 0$, K_{emn} is determined so that $J'_{em}(K_{emn}, \xi_{max}) = 0$,

$$\psi_{omn}(\xi, \eta) = \frac{1}{N_{omn}} S_{om}(K_{omn}, \cos(\eta)) J_{om}(K_{omn}, \cosh(\xi)) \ , \tag{2.16}$$

where the ubscript o stands for odd, N_{omn} is a normalization constant computed so that Eq. (2.5) is satisfied, S_{om} is the odd Mathieu function of the first kind of order m [38], J_{om} is the odd modified Mathieu function of the first kind of order m, $m \in \mathbb{N}, m > 0, n \in \mathbb{N}, n > 0$, K_{omn} is determined so that $J'_{om}(K_{omn}, \xi_{max}) = 0$.

The cut-on frequencies are given by the relations:

$$f_{emn} = \frac{c}{\pi} \sqrt{\frac{K_{emn}}{h^2}} \ , \ f_{omn} = \frac{c}{\pi} \sqrt{\frac{K_{omn}}{h^2}} \ . \tag{2.17}$$

The mode-shapes and the cut-on frequencies of an elliptical cross-section 6.2 cm wide in x_1 and 1.9 cm thick in x_2 are presented for $f_c < 20$ kHz in Fig. 2.4. One can see that with this cross-sectional shape, there are no degenerated modes. The first subscript index corresponds to the number of vertical nodal lines and the second subscript index corresponds to the number of elliptical nodal lines. The odd function features a horizontal straight nodal line.

2.1.1.4 Arbitrary cross-sections

In order to handle geometries with arbitrary cross-sectional shapes, finite differences (FD) are used to solve the Helmholtz equation Eq. (2.1) [39], [40].

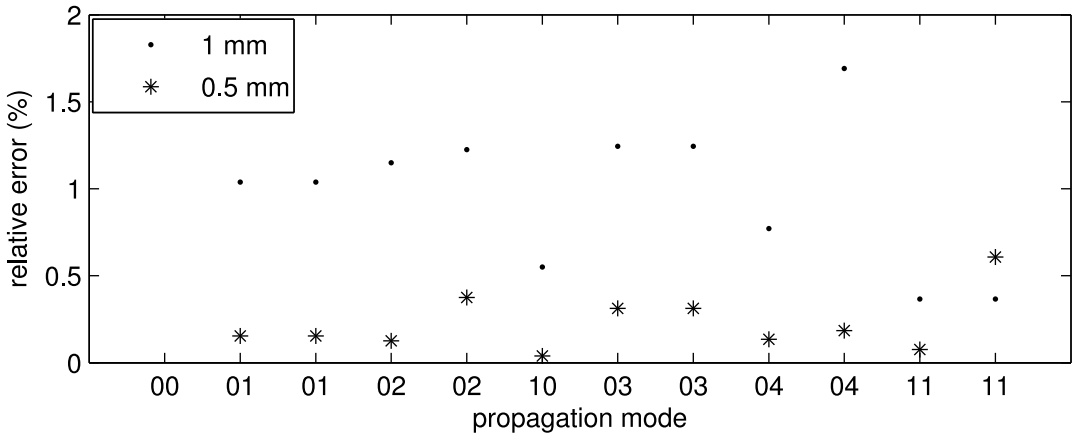


Figure 2.5: Relative error between the cut-on frequencies computed with FD and analytically with Eq. (2.8) for a circular cross-section of diameter 29.5 mm and a temperature of 26.5°C for a spacing of the points of the grid used for FD of 1 mm and 0.5 mm.

A grid of points is generated on the cross-section surface. The spacing between the points of the grid used for the simulations was of 0.5 mm. This has been chosen as a compromise between accuracy and computation time by performing convergence tests and comparing with experimental results and analytical expressions (see section 4.1.2).

In order to test the validity of this approach, it has been applied to circular cross-sections with a diameter of 29.5 mm and compared with the analytical expression of Eq. (2.6). The same mode-shapes as the one presented in Fig. 2.2 are obtained with the FD.

The relative error between the cut-on frequencies computed with FD and analytically with Eq. (2.8) is presented in Fig. 2.5 for a spacing of the points of the grid used for FD of 1 mm and 0.5 mm. For a spacing of 1 mm, the relative error is smaller than 2% and for a spacing of 0.5 mm it is smaller than 1%. Thus, as expected, the accuracy of the computation of the cut-on frequency increases with the resolution. It is interesting to note that the error can be different for two degenerated modes for which the cut-on frequency given analytically is the same. This can be explained by the fact that, due to the angular rotation of these mode-shapes, the distribution of the points of the grid on them is different. The relative error does not necessarily increase with the frequency: the error corresponding to f_{c11}, which is the highest cut-on frequency computed can be smaller than the one of f_{c04}.

2.1.2 Propagation matrices

Once the modal basis corresponding to the cross-section of a waveguide portion has been computed up to a given frequency, it is necessary to compute the variation of the modal amplitudes $\boldsymbol{P}$ and $\boldsymbol{U}$ along the axis x_3.

Solving the Helmholtz equation Eq. (2.1) with the expressions of Eq. (2.4) for the pressure and the velocity allows one to write two relationships between the pressure and velocity modal

amplitudes $\boldsymbol{P}$ and $\boldsymbol{U}$ at two different locations. Consider two points located within the waveguide at the abscissas $x_3^{(0)}$ and $x_3^{(1)}$. $\boldsymbol{P}^{(0)}$ and $\boldsymbol{U}^{(0)}$ at $x_3^{(0)}$ can be known from $\boldsymbol{P}^{(1)}$ and $\boldsymbol{U}^{(1)}$ at $x_3^{(1)}$ with

$$\boldsymbol{P}^{(0)} = \boldsymbol{D_1}\boldsymbol{P}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{U}^{(1)}, \tag{2.18}$$

$$\boldsymbol{U}^{(0)} = \boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\boldsymbol{P}^{(1)} + \boldsymbol{D_1}\boldsymbol{U}^{(1)}, \tag{2.19}$$

where $\boldsymbol{D_1} = \mathrm{diag}(\cos(k_n d))$ is a diagonal matrix whose terms are $\cos(k_n d)$ with $d = x_3^{(1)} - x_3^{(0)}$, $\boldsymbol{D_2} = \mathrm{diag}(j\sin(k_n d))$ and $\boldsymbol{Z_c}$ is the characteristic impedance matrix obtained as $\boldsymbol{Z_c} = \mathrm{diag}\left(\frac{k\rho c}{k_n S}\right)$.

Using $\boldsymbol{P}^{(i)} = \boldsymbol{Z}^{(i)}\boldsymbol{U}^{(i)}$, with $i = 0$ or $i = 1$, allows one to write a relationship between the impedance matrices $\boldsymbol{Z}^{(0)}$ and $\boldsymbol{Z}^{(1)}$ at $x_3^{(0)}$ and $x_3^{(1)}$

$$\boldsymbol{Z}^{(0)} = \left[\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c}\right]\left[\boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_1}\right]^{-1} . \tag{2.20}$$

Following the same procedure and using the equation $\boldsymbol{U}^{(i)} = \boldsymbol{Y}^{(i)}\boldsymbol{P}^{(i)}$ allows one to write three other relationships between the impedance and the admittance matrices at $x_3^{(0)}$ and $x_3^{(1)}$

$$\boldsymbol{Z}^{(0)} = \left[\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}\right]\left[\boldsymbol{D_2}\boldsymbol{Z_c}^{-1} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)}\right]^{-1}, \tag{2.21}$$

$$\boldsymbol{Y}^{(0)} = \left[\boldsymbol{D_1}\boldsymbol{Y}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\right]\left[\boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)} + \boldsymbol{D_1}\right]^{-1}, \tag{2.22}$$

$$\boldsymbol{Y}^{(0)} = \left[\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)}\right]\left[\boldsymbol{D_2}\boldsymbol{Z_c} + \boldsymbol{D_1}\boldsymbol{Z}^{(1)}\right]^{-1}. \tag{2.23}$$

However $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ in Eqs. (2.20), (2.21), (2.22) and (2.23) contain very large terms for evanescent modes which induce round off numerical errors. To ensure numerical stability a new diagonal matrix $\boldsymbol{D_3} = \mathrm{diag}(\tan(k_n d))$ is introduced. Equations (2.20), (2.21), (2.22) and (2.23) are rewritten using only $\boldsymbol{D_3}^{-1}$ and $\boldsymbol{D_2}^{-1}$ instead of $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$,

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}[\boldsymbol{Z}^{(1)} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}]\boldsymbol{D_2}^{-1}\boldsymbol{Z_c} , \tag{2.24}$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}[\boldsymbol{I} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Y}^{(1)}\boldsymbol{Z_c}]\boldsymbol{D_2}^{-1}\boldsymbol{Z_c} , \tag{2.25}$$

$$\boldsymbol{Y}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}[\boldsymbol{Y}^{(1)} + (j\boldsymbol{D_3})^{-1}]\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1} , \tag{2.26}$$

$$\boldsymbol{Y}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}[\boldsymbol{I} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)}]\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1} , \tag{2.27}$$

$\boldsymbol{I}$ being the identity matrix. For a detailed computation of these four equations, see appendix C.

2.1.3 Radiation impedance

Once relations between the modal amplitudes, impedance and admittance matrices at two different abscissas of a constant cross-section waveguide have been derived, it is necessary to have a boundary condition in order to compute a solution for Eq. (2.1). At the open end of

a waveguide, the boundary condition can be described with a radiation impedance matrix. It can be obtained analytically in the case of circular [41] or rectangular [30], [42] cross-section. For the intermediate cross-sections, the boundary condition is the modal amplitudes, the impedance or the admittance matrix at the ends of the neighboring sections.

2.1.3.1 Numerical computation of the radiation impedance matrix

In the case of an arbitrary cross-section, no simple analytical expression of the radiation impedance matrix can be derived. However it can be obtained by numerically integrating the expression

$$Z_{mn} = \frac{j\omega\rho}{2\pi S^2} \int_S \int_S \psi_m(x_1', x_2')\psi_n(x_1, x_2)\frac{e^{-jkh}}{h} dS dS \ , \tag{2.28}$$

with $h = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2}$ and S being the cross-section surface at the open end of the waveguide.

This can be achieved by computing the values of ψ_m and ψ_n on each point of the grid used for the FD and approximating the integrals of Eq. (2.28) by finite summations

$$Z_{mn} = \frac{j\omega\rho}{2\pi N^2} \sum_{a=1}^{N} \sum_{b=1}^{N} \psi_m(x_{1a}', x_{2a}')\psi_n(x_{1b}, x_{2b})\frac{e^{-jkh_{ab}}}{h_{ab}} \ , \tag{2.29}$$

with $h_{ab} = \sqrt{(x_{1b} - x_{1a}')^2 + (x_{2b} - x_{2a}')^2}$. Note that one does not divide by the area S^2, since the area elements dS have not been taken into account in the summation above. On the other hand, one must divide by the number of points N^2.

A particular case appears when $h = 0$: infinite values are generated and one must avoid this case if Eq. (2.29) is used, and the accuracy of the computation is affected. In order to work around this limitation, the position of the points on which the amplitude of one of the eigen-modes is computed is expressed in polar coordinates (r, θ). The origin of the polar landmark is set on the point whose position is defined in Cartesian coordinates so that the radial coordinate is equal to h. Thus, the origin is moved for each point expressed in Cartesian coordinates. In this case, the infinitesimal area element is expressed as $dS = hdrd\theta$ in the polar landmark, and Eq. (2.28) can be rewritten

$$Z_{mn} = \frac{j\omega\rho}{2\pi S^2} \int_S \int_S \psi_m(r, \theta)\psi_n(x_1, x_2)e^{-jkh} dr d\theta dx_1 dx_2 \ , \tag{2.30}$$

and the division by h which induces the singularity in Eq. (2.29) is removed. Because a regular discretization over x_1 and x_2 induces an irregular discretization over r and θ, it is necessary to generate a second grid of N_p points regularly spaced over r and θ. This insures that the area element $hdrd\theta$ remains similar for each point (r_b,θ_b) and that the simplification introduced is valid. As a consequence, a new grid is generated for each point (x_{1a},x_{2a}). Eq. (2.30) can then be discretized the following way:

$$Z_{m,n} = \frac{j\omega\rho}{2\pi N} \sum_{a=1}^{N} \frac{1}{\sum_{b=1}^{N_p} h_{ab}} \sum_{b=1}^{N_p} \psi_m(r_b, \theta_b)\psi_n(x_{1a}, x_{2a})e^{-jkh_{ab}} \ . \tag{2.31}$$

As for Eq. (2.29), one does not divide by the square of the area, since the area elements dS are not taken into account. However, one must divide the polar coordinate summation by $\sum_{b=1}^{N_p} h_{ab}$ for each value of a because it is the equivalent of $\int_S r dr d\theta$. The overall result needs also to be divided by N.

2.1.3.2 Validation of the numerical computation of the impedance matrix with analytical expressions

In order to validate the numerical method described earlier, its outcomes have been compared to the values obtained with the analytical expressions given by [41] and [30] which have been implemented in Matlab/Octave. Note that when implementing Eqs. (3.40) and (3.35) of [30], one must be careful that the sine cardinal function is defined as $\mathrm{sinc}(x) = \frac{\sin(x)}{x}$ in this document, whereas it is defined as $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$ in Matlab and Octave.

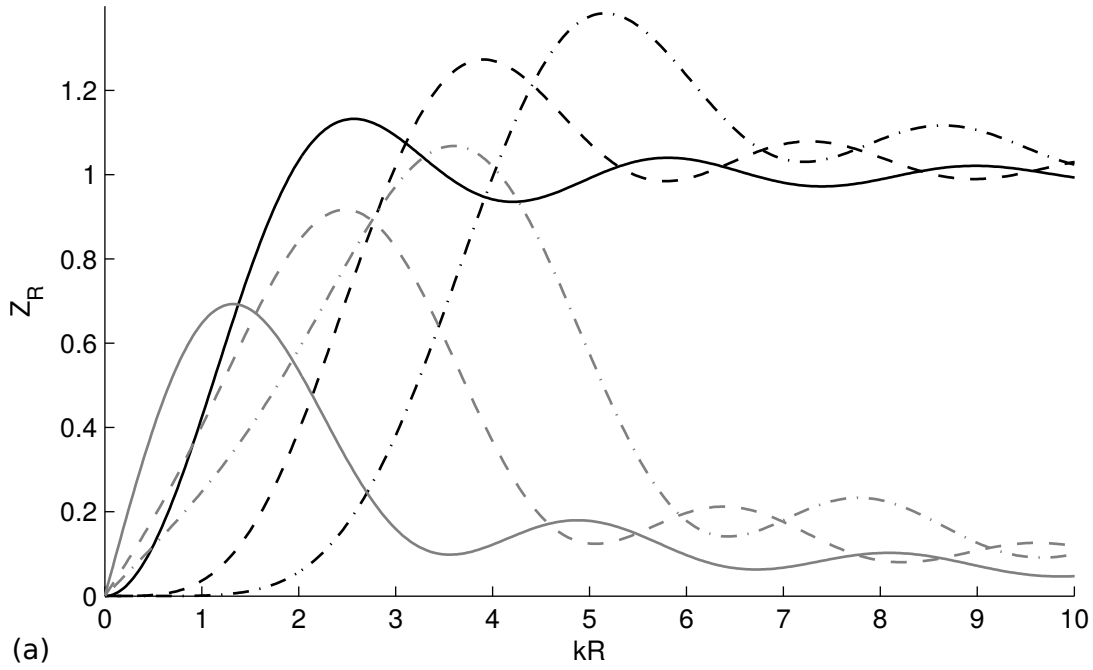
The radiation impedance matrix $\boldsymbol{Z_R}$ has been computed for a circular and a square opening measuring 29.5 mm in the diameter and the width respectively. The data computed with the numerical method and the expression given by [30] have been normalized by $\frac{\rho c}{S}$, S being the surface of the exit. This normalization is already applied in the expression given by [41]. In order to avoid any disturbance due to the fact that the mode-shapes obtained with FD can potentially be slightly different from the analytical ones, Eqs. (2.6) and (2.9) have been used to compute the mode-shapes for the numerical approach. A resolution of 0.5 mm has been used. This corresponds to a ratio of resolution over largest dimension of about 1/60.

The real and imaginary parts of the first terms of the circular cross-section are presented in Figs. 2.6a and 2.7a for the diagonal and coupling terms respectively. For the square cross-section, the first terms of $\boldsymbol{Z_R}$ are presented in Figs. 2.8a and 2.9a for the diagonal and the coupling terms respectively. Note that Eq. 3.41 of [30] gives radiation impedance matrix terms only for the modes without nodal lines in the diameter (which pass by the center of the cross-section, see the modes ψ_{00}, ψ_{02}, ψ_{20} and ψ_{22} in Fig. 2.3). Thus, it is only possible to compare the numerical and analytical method for modes with even indices: the indices of Eq. (2.31) are two times greater than the one of Eq. 3.41 of [30].

The real part of the radiation impedance terms characterizes the losses by radiation at the open end of the waveguide. The greater it is, the greater are the losses. The imaginary part part characterizes the inertia introduced by the opening which varies the phase shift introduced to the wave reflected back to the waveguide. The greater it is, the greater is the phase shift increase. In the plane wave theory, it can be accounted for with a length correction.

The same curves as the one reported in the literature [30], [41] are observed in Figs. 2.6a, 2.7a, 2.8a and 2.9a. Close to $kR = 0$, the radiation impedance tends to 0 for all terms of the matrix, which corresponds to the perfect open end boundary condition.

In the case of the diagonal terms (see Figs. 2.6a and 2.8a), the real part increases and then tends to one and the imaginary part increases and then decreases again to tend to zero. Thus, it tends to the free field plane wave impedance. It corresponds to the fact that at high enough



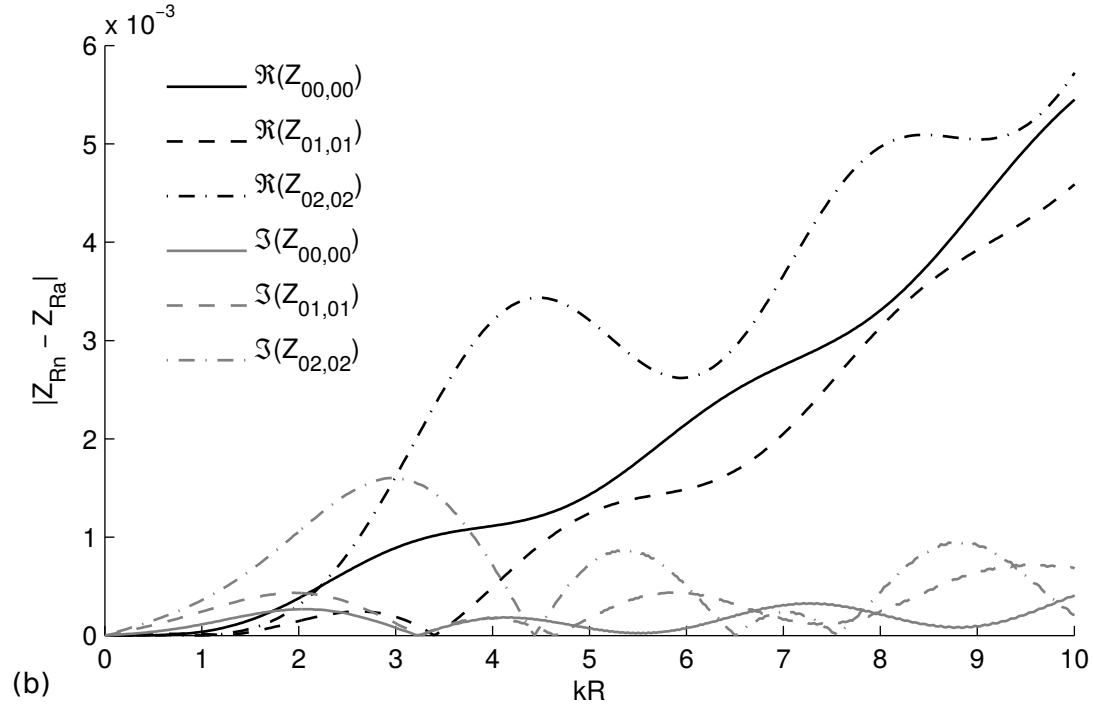
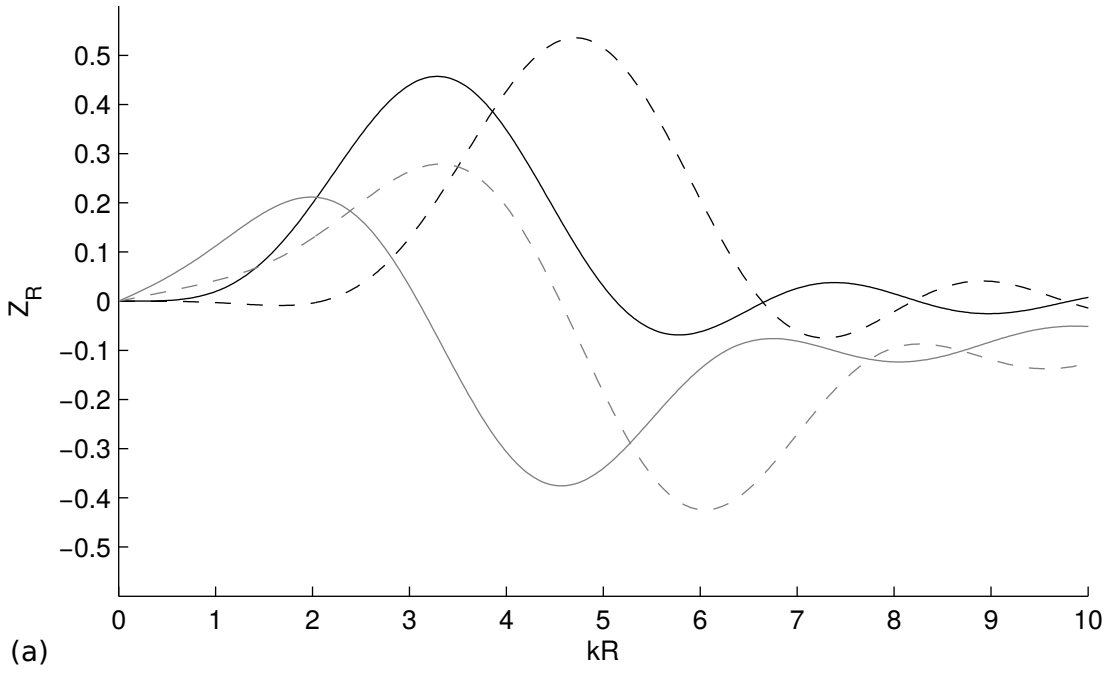


Figure 2.6: First diagonal terms of the normalized radiation impedance matrix $\boldsymbol{Z_R}$ of a circular opening of 29.5 mm in diameter computed with the expression provided by Zorumski [41] ($\boldsymbol{Z_{Ra}}$) and numerically ($\boldsymbol{Z_{Rn}}$, Eq. (2.31)), (a) average value given by both methods, (b) difference between both methods.



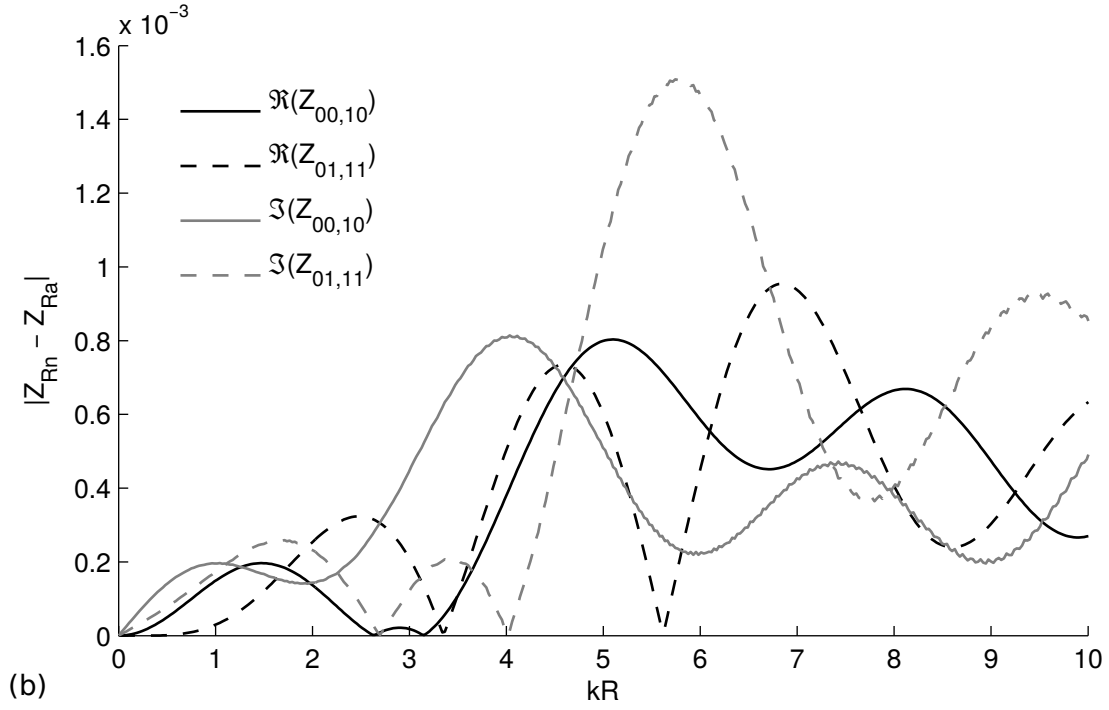
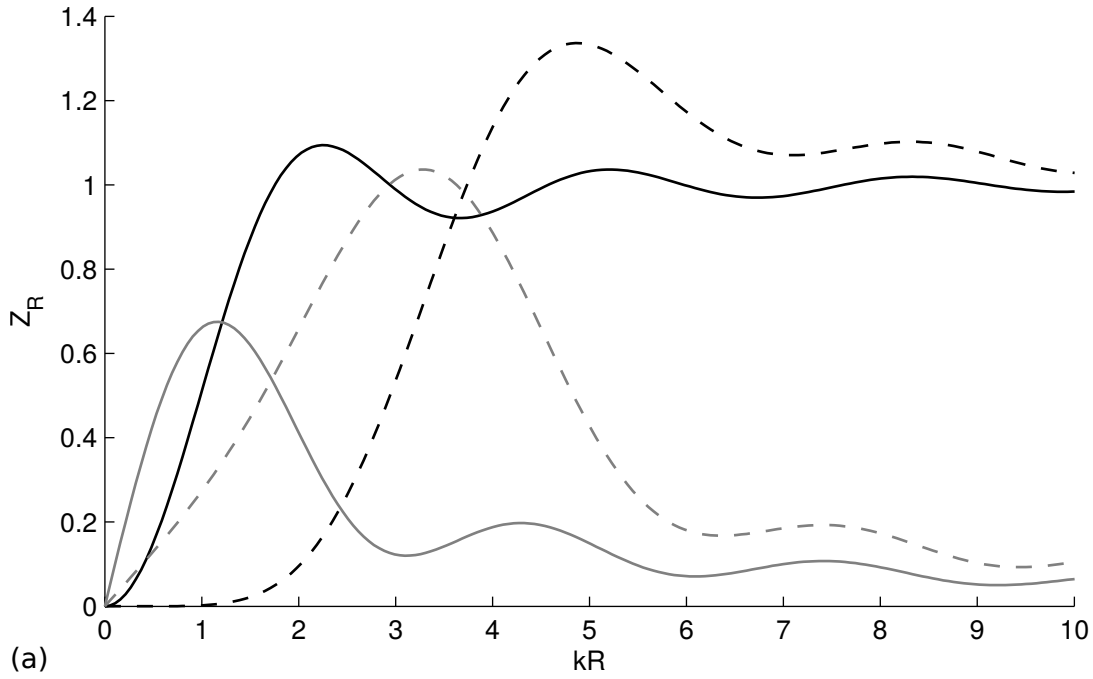


Figure 2.7: First coupling terms of the normalized radiation impedance matrix $\boldsymbol{Z_R}$ of a circular opening of 29.5 mm in diameter computed with the expression provided by Zorumski [41] ($\boldsymbol{Z_{Ra}}$) and numerically ($\boldsymbol{Z_{Rn}}$, Eq. (2.31)), (a) average value given by both methods, (b) difference between both methods.



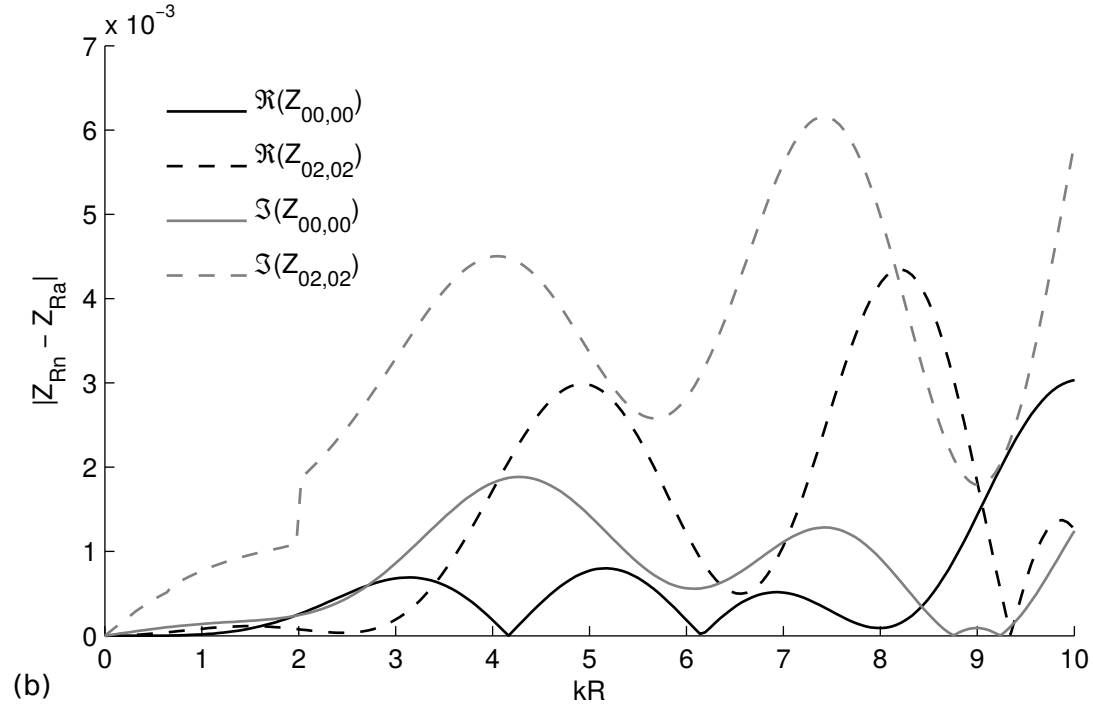
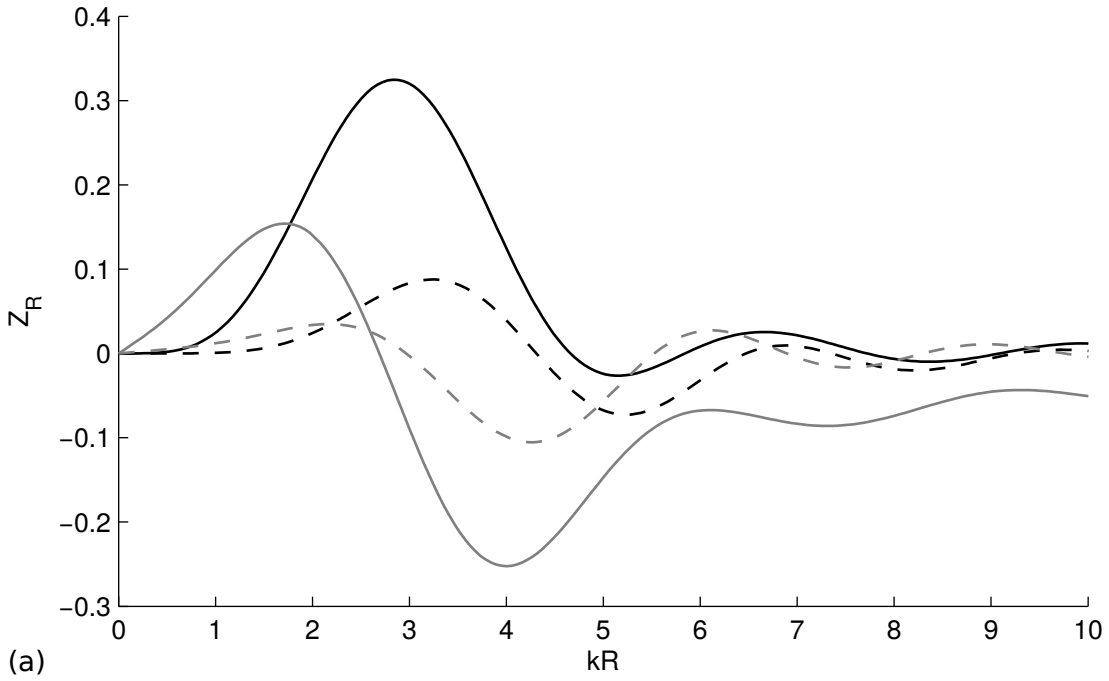


Figure 2.8: First diagonal terms of the normalized radiation impedance matrix $\boldsymbol{Z_R}$ of a square opening having a width of 29.5 mm computed with the expression provided by Kemp [30] ($\boldsymbol{Z_{Ra}}$) and numerically ($\boldsymbol{Z_{Rn}}$, Eq. (2.31)), (a) average value given by both methods, (b) difference between both methods.



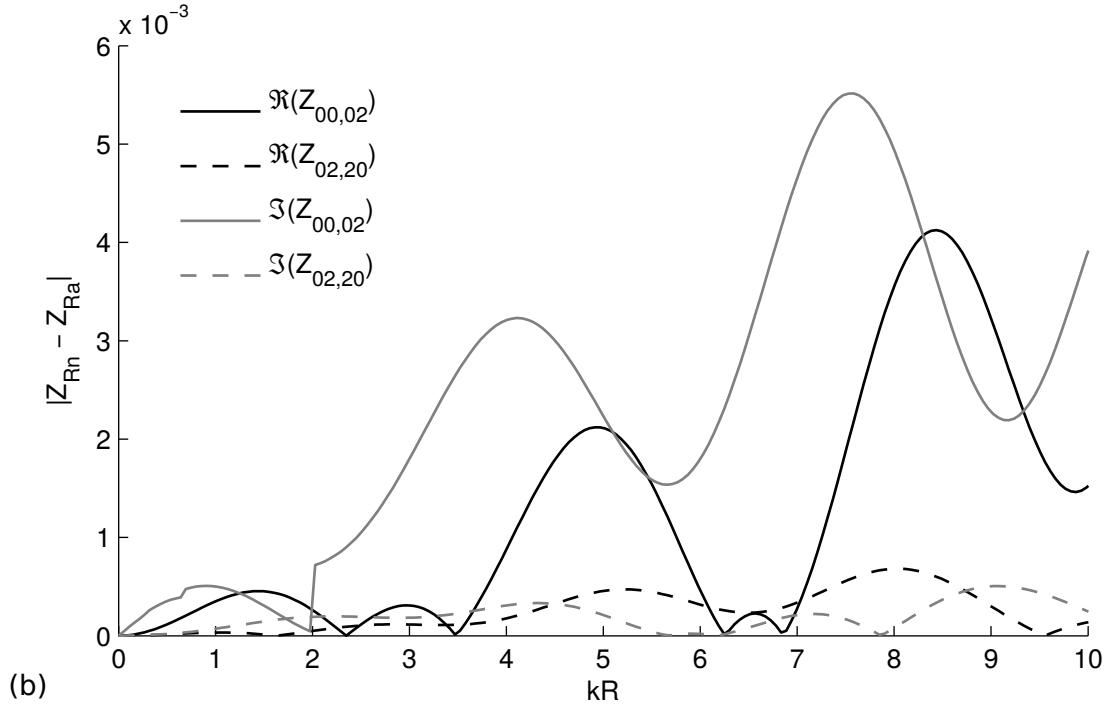


Figure 2.9: First coupling terms of the normalized radiation impedance matrix $\boldsymbol{Z_R}$ of a square opening having a width of 29.5 mm computed with the expression provided by Kemp [30] ($\boldsymbol{Z_{Ra}}$) and numerically ($\boldsymbol{Z_{Rn}}$, Eq. (2.31)), (a) average value given by both methods, (b) difference between both methods.

frequencies the free field assumption holds and there is no more the effect of the boundary condition.

The coupling terms quantify how the different propagation modes are coupled at the open end of the waveguide (see Figs. 2.7a and 2.9a). Thus, even if not excited at the entrance, a HOM coupled with another one which is excited can be excited by coupling and propagate if the frequency is greater than its cut-on frequency. Their real and imaginary parts tends to zero close to $kR = 0$, then increases and oscillates between negative and positive values, and decreases to tend to zero at high frequency. This corresponds to the fact that at a high enough frequency the free field assumption holds and no coupling can be possible. In the case of a circular opening of 29.5 mm and considering only the modes propagating under 20 kHz only the coupling terms $Z_{00,10}$, $Z_{01,11}$, $Z_{10,00}$ and $Z_{11,01}$ are non zero.

The radiation impedance matrix obtained with the numerical approach $\boldsymbol{Z_{Rn}}$ has been compared with the one obtained with the analytical method $\boldsymbol{Z_{Ra}}$. The absolute difference between the real part and the imaginary part $|\boldsymbol{Z_{Rn}} - \boldsymbol{Z_{Ra}}|$ is presented for the circular opening in Figs. 2.6b and 2.7b for the diagonal and the coupling terms respectively. For the square opening, $|\boldsymbol{Z_{Rn}} - \boldsymbol{Z_{Ra}}|$ is presented in Figs. 2.8b and 2.9b for the diagonal and the coupling terms respectively. In this case the relative error is not meaningful to compute everywhere because when the curves cross zero, unrealistically large terms are computed.

The accuracy of the numerical computation is quite good: the absolute difference between both methods is smaller than 7.10^{-3}, which is three orders of magnitude smaller than the real and the imaginary parts of the impedance. For the maxima of diagonal terms, this yields a relative error smaller than 1% . The error tends to be smaller for the coupling terms. For a given term, it varies with maxima and minima, but globally it tend to increase with kR.

2.2 Junction between waveguide portions

Once a solution of the Helmholtz equation (Eq. 2.1) can be derived for each section of the waveguide, it is necessary to join the different sections in order have a solution for the whole waveguide.

2.2.1 Projection matrix

At a junction between two ducts with different cross-sections it is necessary to find a relationship between the modal basis of each section. Let us consider two ducts a and b whose cross-sectional surfaces are S_a and S_b respectively (see Fig. 2.10). Applying the continuity of both the pressure p and the volume velocity u_{x_3} on each side of the junction yields

$$\boldsymbol{\psi_a^t P_a} = \boldsymbol{\psi_b^t P_b} \, , \tag{2.32}$$

$$\frac{1}{S_a}\boldsymbol{\psi_a^t U_a} = \frac{1}{S_b}\boldsymbol{\psi_b^t U_b} \, . \tag{2.33}$$

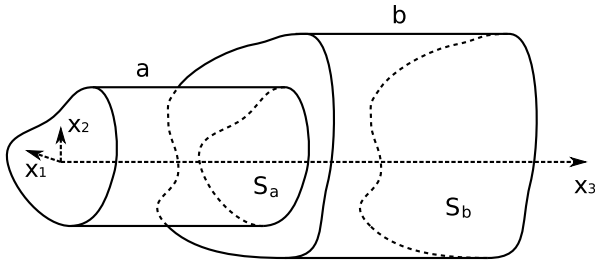


Figure 2.10: Junction between two ducts of arbitrary cross-section shape.

Hereafter the subscript b always refers to the duct with the largest cross-section and the subscript a to the smallest one. If one assumes that $S_a < S_b$ and S_a lies inside S_b, multiplying both sides of Eq. (2.32) with $\boldsymbol{\psi}_{\boldsymbol{a}}^{*}/S_a$ and integrating over S_a leads to

$$\boldsymbol{P_a} = \frac{1}{S_a}\left[\int_{S_a} \boldsymbol{\psi}_{\boldsymbol{a}}^{*}\boldsymbol{\psi}_{\boldsymbol{b}}^{t} dS_a\right]\boldsymbol{P_b}\ . \tag{2.34}$$

In the same way multiplying Eq. (2.33) with $\boldsymbol{\psi}_{\boldsymbol{b}}^{*}$ and integrating over S_b yields

$$\boldsymbol{U_b} = \frac{1}{S_a}\left[\int_{S_b} \boldsymbol{\psi}_{\boldsymbol{b}}^{*}\boldsymbol{\psi}_{\boldsymbol{a}}^{t} dS_b\right]\boldsymbol{U_a}\ . \tag{2.35}$$

Yet we have $\boldsymbol{\psi_a} = 0$ on $S_b - S_a$ so

$$\int_{S_b} \boldsymbol{\psi}_{\boldsymbol{b}}^{*}\boldsymbol{\psi}_{\boldsymbol{a}}^{t} dS_b = \int_{S_a} \boldsymbol{\psi}_{\boldsymbol{b}}^{*}\boldsymbol{\psi}_{\boldsymbol{a}}^{t} dS_a = \left[\int_{S_a} \boldsymbol{\psi}_{\boldsymbol{a}}^{*}\boldsymbol{\psi}_{\boldsymbol{b}}^{t} dS_a\right]^{\dagger}\ ,$$

where the symbol † means transpose conjugate. Thus, defining the projection matrix $\boldsymbol{F}$ as

$$\boldsymbol{F} = \frac{1}{S_a}\int_{S_a} \boldsymbol{\psi}_{\boldsymbol{a}}^{*}\boldsymbol{\psi}_{\boldsymbol{b}}^{t} dS, \tag{2.36}$$

equations (2.34) and (2.35) can be rewriten as

$$\boldsymbol{P_a} = \boldsymbol{F}\boldsymbol{P_b}\ , \tag{2.37}$$

$$\boldsymbol{U_b} = \boldsymbol{F}^{\dagger}\boldsymbol{U_a}\ . \tag{2.38}$$

Given that $\boldsymbol{P} = \boldsymbol{Z}\boldsymbol{U}$ and $\boldsymbol{U} = \boldsymbol{Y}\boldsymbol{P}$, Eq. (2.37) and (2.38) can be expressed as

$$\boldsymbol{Z_a} = \boldsymbol{F}\boldsymbol{Z_b}\boldsymbol{F}^{\dagger}\ , \tag{2.39}$$

$$\boldsymbol{Y_b} = \boldsymbol{F}^{\dagger}\boldsymbol{Y_a}\boldsymbol{F}\ . \tag{2.40}$$

An analytic expression of F_{mn} can be found in the particular case of two circular cross-sections sharing the same central axis (see [30]):

$$F_{m0a,m0b} = \frac{2\beta\gamma_{m0b}J_1(\beta\gamma_{m0b})}{(\beta^2\gamma_{m0b}^2 - \gamma_{m0a}^2)J_0(\gamma_{m0b})}\ , \tag{2.41}$$

where $\beta = R_a/R_b$ is the ratio of the radii of both cross-sections and γ_{m0} is the order m zero of the derivative of the Bessel function of first kind of order zero J_0. Note that, in this case, only the mode-shapes without radial nodal lines can generate non-zero values of the $\boldsymbol{F}$ matrix.

It is also possible to find analytic expressions for other particular cases such as junctions between rectangular cross-sections or axisymmetric elliptical cross-sections. However, it is difficult to find an analytic expression for the case of an eccentric junction between circular and elliptical cross-sections, and it is not possible to solve analytically the cases with arbitrary cross-sections. Thus, numerical integration has been used to solve these problems. On the other hand, this allows one to handle the case of the junction between two waveguide portions having different cross-sectional shape.

The value of ψ_a and ψ_b are evaluated at each point of the grid used for the FD computation of the propagation modes. The coupling coefficient can then be evaluated by computing the sum

$$F_{a,b} = \frac{1}{N} \sum_{i=1}^{N_p} \psi_a^*(x_{1i}, x_{2i}) \psi_b(x_{1i}, x_{2i}). \tag{2.42}$$

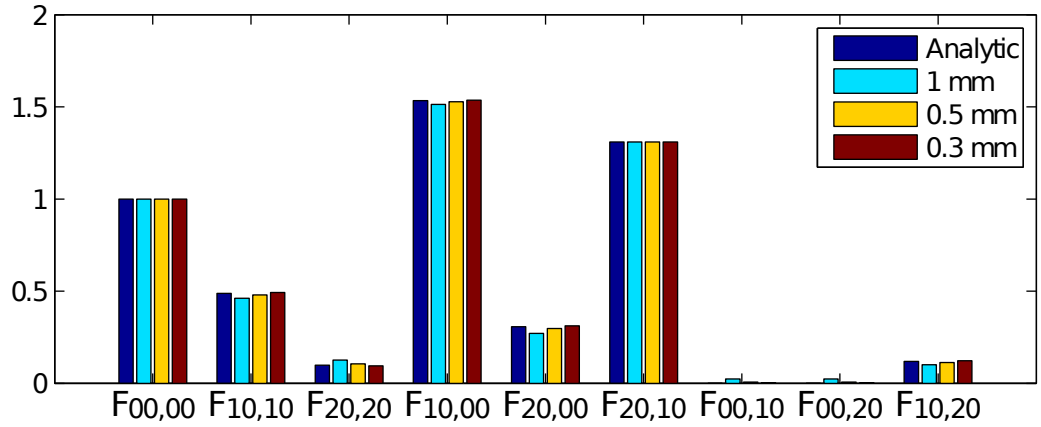
Note that the case of a junction whose cross-sections are not enclosed in each other can be solved by inserting an intermediate zero-length waveguide portion whose cross-section is the intersection of the original two neighboring cross-sections.

2.2.2 Validation against analytical case

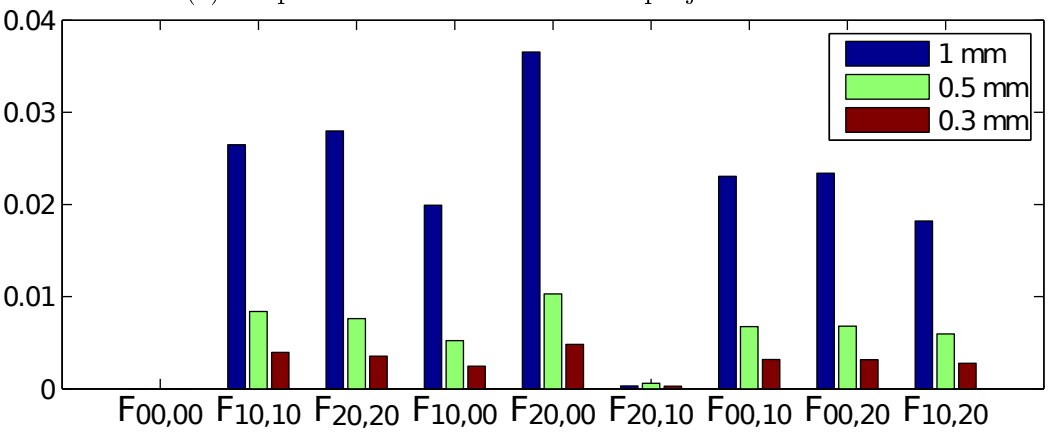
In order to test the validity and the accuracy of this approach, it has been compared with the values given by Eq. (2.41). The case of a junction with radii equal to 14.5 mm and 29.5 mm (this corresponds to a geometry used in what is presented later, see section 3.1) have been solved with both methods. The values of the first non-zeros terms of the matrix $\boldsymbol{F}$ are presented in Fig. 2.11. Note that the comparison is made on the amplitude of the terms. Sign difference can occur between analytic and numeric computations. However, this does not affect the overall outcome of the simulation.

As expected the agreement improves when the resolution is increased. The maximal amplitude difference is of 0.037 with 1 mm, 0.011 with 0.5 mm and 0.005 with 0.3 mm. The relative amplitude difference can reach 29% with 1 mm, 8% with 0.5 mm and 4% with 0.3 mm. The error in not the same for the different terms of matrix $\boldsymbol{F}$. The numerical solution is exact for $F_{00,00}$ and thus there is no amplitude difference with the analytical solution. The amplitude difference is maximal for $F_{20,00}$ (0.0365) and minimal for $F_{20,10}$ (0.0007). The maximum of relative amplitude difference is reached with $F_{20,20}$ (28.6 %) and the minimum with $F_{20,10}$ (0.05%). Computing relative amplitude difference for $F_{00,10}$ and $F_{00,20}$ is not meaningful because the analytical value is zero for those terms.

The maximal relative amplitude difference obtained with the resolution used for the simulations, 0.5 mm, can appear as high (8%). However, it has been chosen as a compromise



(a) Amplitude of some terms of the projection matrix



(b) Difference between the analytical and the numerical computations

Figure 2.11: Amplitude of some terms of the projection matrix $\boldsymbol{F}$ computed analytically (with Eq. (2.41)) and numerically (with Eq. (2.42)) with different resolutions for the case of a junction between two concentric cross-sections of radii 14.5 mm and 29.5 mm and differences between the analytical and the numerical computations.

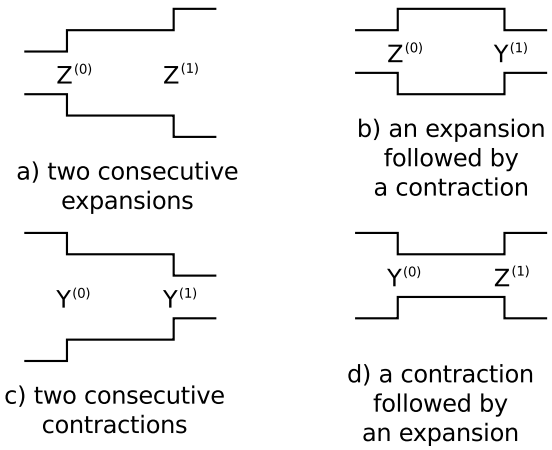


Figure 2.12: The four different configurations for backward propagation of the impedance and admittance mattrices.

between accuracy and computation time, and the global convergence tests showed that the overall convergence is satisfying (see section 4.1.2).

2.2.3 The four types of possible junctions

The impedance matrix can be back propagated towards the entrance of the waveguide. To do so, Eqs. (2.39) and (2.40) allow one to propagate the impedance and admittance matrices through a junction. However these matrices can be ill conditioned and inversion must be avoided. Thus the impedance matrix must be used for expansions and the admittance matrix for contractions. This constraint introduces four different cases for the projection of the impedance or admittance matrices from a junction to another which are described in Fig. 2.12.

For case a) both junctions are expanding so the impedance matrices can be propagated through the junctions with Eq. (2.39) and $\boldsymbol{Z}^{(0)}$ can be computed from $\boldsymbol{Z}^{(1)}$ with Eq. (2.24). On the contrary in case c) both junctions are contracting so the admittance matrices can be propagated through the junctions with Eq. (2.40) and $\boldsymbol{Y}^{(0)}$ can be computed from $\boldsymbol{Y}^{(1)}$ with Eq. (2.26). The case b) is expanding and then contracting so the admittance matrix can be propagated through the junction of the right and the impedance matrix can be propagated through the junction of the left. The impedance $\boldsymbol{Z}^{(0)}$ can be computed from the admittance $\boldsymbol{Y}^{(1)}$ with Eq. (2.25). On the opposite the case d) is contracting and then expanding so the impedance matrix can be propagated through the junction of the right and the admittance matrix can be propagated through the junction of the left. The admittance $\boldsymbol{Y}^{(0)}$ can be computed from the impedance $\boldsymbol{Z}^{(1)}$ with Eq. (2.27).

The problem of a junction whose cross-sections are not enclosed in each other can be solved by inserting an intermediate zero-length waveguide portion whose cross-section is the intersection of the original two neighboring cross-sections.

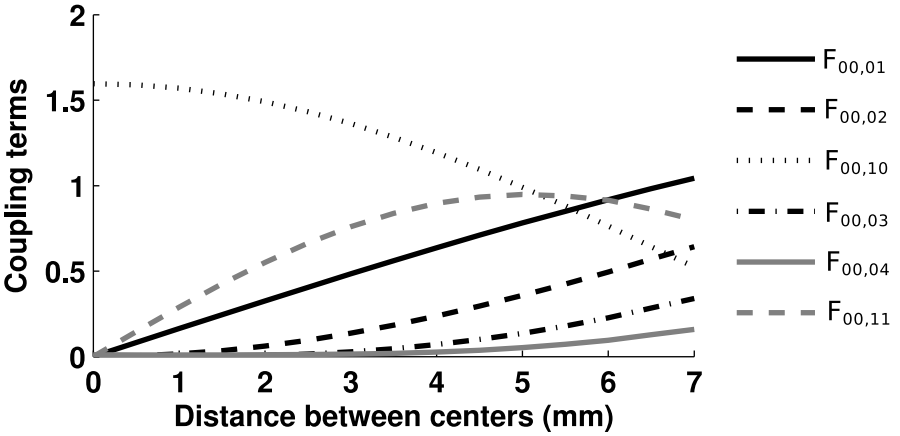


Figure 2.13: Coupling terms of the projection matrix $\boldsymbol{F}$ of Eq. (2.36) between the plane mode ψ_{00} and the HOM whose cut-on frequency is below 20 kHz (ψ_{01}, ψ_{02}, ψ_{10}, ψ_{03}, ψ_{04} and ψ_{11}) for a junction between two cylindrical sections of diameter 14.5 mm and 29.5 mm with a distance between their centers varying from 0 to 7 mm. The amplitude of the terms $F_{00,01}$, $F_{00,02}$, $F_{00,03}$, $F_{00,04}$ and $F_{00,11}$, correspond to the combination of the amplitude of both degenerated HOM ψ_{mn1} and ψ_{mn2}: $F_{00,mn} = \sqrt{F_{00,mn1}^2 + F_{00,mn2}^2}$.

2.2.4 Effect of eccentricity on coupling terms

In order to investigate the effect of the degree of eccentricity of a junction between two cylindrical waveguide portions on the excitation of HOM, the projection matrix $\boldsymbol{F}$ has been computed by varying the distance between the centers of two circular cross-sections of diameter 14.5 mm and 29.5 mm from 0 mm to 7 mm by step of 0.5 mm. This corresponds to the progressive transition from a concentric configuration to a fully eccentric one.

The coupling terms of the projection matrix $\boldsymbol{F}$ between the plane mode ψ_{00} and the HOM whose cut-on frequency is below 20 kHz (see Fig. 2.2) are presented in Fig. 2.13. The amplitude of the terms $F_{00,01}$, $F_{00,02}$, $F_{00,03}$, $F_{00,04}$ and $F_{00,11}$, corresponds to the combination of the amplitude of both degenerated HOM ψ_{mn1} and ψ_{mn2}: $F_{00,mn} = \sqrt{F_{00,mn1}^2 + F_{00,mn2}^2}$.

The coupling terms between the plane mode ψ_{00} and the modes with nodal lines on the diameter (ψ_{01}, ψ_{02}, ψ_{03}, ψ_{04} and ψ_{11}, see Fig. 2.6) are 0 when there is no shifting between the centers of the sections (which corresponds to the concentric configuration). Indeed, theoretically in Eq. (2.36) the contributions of the plane mode ψ_{00} on either side of the nodal lines of these modes compensate each other exactly.

When the eccentricity (the distance between the centers) is increased, the excitation of the modes featuring nodal lines on the diameter is increased. On the contrary the excitation of the axisymmetric mode ψ_{10} is decreased when the junction becomes more eccentric. This is due to the fact that the circular nodal line of this mode gets closer to the middle of the smallest cross-section. Thus the contributions of the plane mode ψ_{00} to excite ψ_{10} tend to

annihilate.

2.3 Computation of pressure and volume velocity

2.3.1 Internal field

Once the impedance and or the admittance matrices $\boldsymbol{Z}$ and $\boldsymbol{Y}$ have been computed for each section of the waveguide, it is possible to compute the modal amplitudes of the pressure and the volume velocity $\boldsymbol{P}$ and $\boldsymbol{U}$ at the entrance and to propagate it up to the exit.

The pressure or the volume velocity of the source can be projected on the modal basis which describes the pressure and velocity field at the source location. For a source delivering a velocity v_s on a surface S_s the modal amplitude $\boldsymbol{U_s}$ corresponding to the source can be known with

$$\boldsymbol{U_s} = \int_{S_s} \boldsymbol{\psi}^* v_s dS_s \ . \tag{2.43}$$

The pressure and the velocity modal amplitudes can then be propagated from the entrance to the open end. Introducing $\boldsymbol{Y^{(1)}}$ and $\boldsymbol{Z^{(1)}}$ into Eqs. (2.18) and (2.19) leads to

$$\boldsymbol{P}^{(1)} = \left[\boldsymbol{D_1} + \boldsymbol{D_2 Z_c Y^{(1)}}\right]^{-1} \boldsymbol{P}^{(0)} \ , \tag{2.44}$$

$$\boldsymbol{U}^{(1)} = \left[\boldsymbol{D_2 Z_c^{-1} Z^{(1)}} + \boldsymbol{D_1}\right]^{-1} \boldsymbol{U}^{(0)} \ , \tag{2.45}$$

which allows one to propagate $\boldsymbol{P}$ or $\boldsymbol{U}$ (depending if the impedance or the admittance is known) through a constant cross-section waveguide portion. The vectors $\boldsymbol{P}$ and $\boldsymbol{U}$ can then be propagated through the junctions using Eqs. (2.37) and (2.38). Equations (2.18) and (2.19) allow one to compute $\boldsymbol{P}$ and $\boldsymbol{U}$ at any abscissa x_3 and the pressure p and the axial velocity u_{x_3} can be computed with Eq. (2.2) at any point inside the waveguide.

2.3.2 Radiated acoustic pressure

2.3.2.1 Free field

The pressure field radiated by the waveguide can be computed using the Rayleigh-Sommerfield integral

$$p(x_1, x_2, x_3) = \sum_{n=0}^{\infty} \frac{j\omega\rho}{2\pi S} \int_S U_n \psi_n(x_1', x_2') \frac{e^{jkh}}{h} dS, \tag{2.46}$$

with U_n the modal amplitude of the particle velocity, $h = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2}$, S being the surface at the open end of the waveguide and (x_1', x_2', x_3') the coordinates of the points on S.

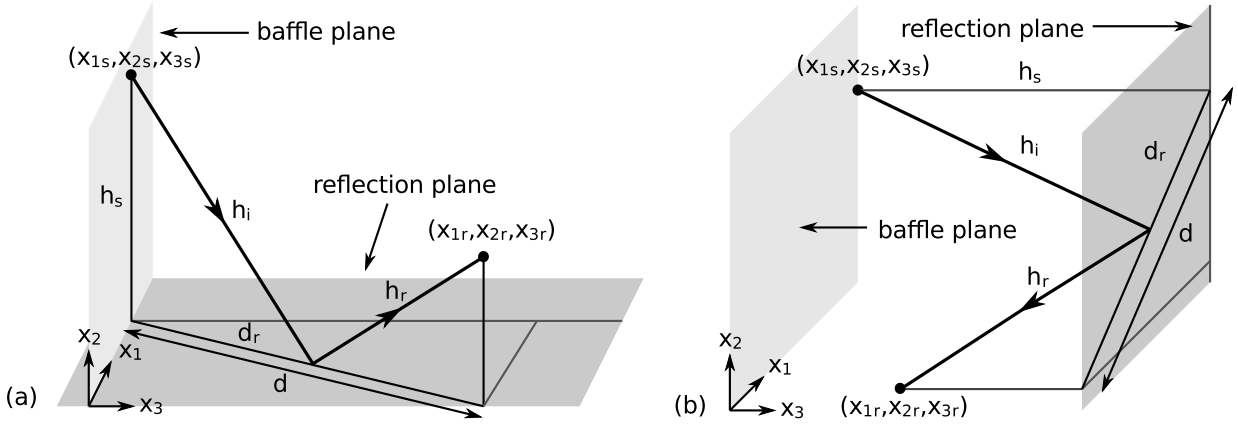


Figure 2.14: Path followed by a reflected sound wave emitted by a baffled vocal tract replica. (a) Reflection on a plane perpendicular to the baffle (b) Reflection on a plane parallel to the baffle located behind the reception point

This integral can be approximated considering a finite summation over $N+1$ modes and replacing the integral by a finite summation evaluated on a grid of N_p points

$$P(x_1, x_2, x_3, k) = \sum_{n=0}^{N} \frac{j\omega\rho}{2\pi N_p} \sum_{i=1}^{N_p} U_n \psi_n(x'_{1i}, x'_{2i}) \frac{e^{jkh_i}}{h_i}, \tag{2.47}$$

with $h_i = \sqrt{(x_1 - x'_{1i})^2 + (x_2 - x'_{2i})^2 + (x_3 - x'_{3i})^2}$, $(x'_{1i}, x'_{2i}, x'_{3i})$ being the coordinates of the points of the grid. Note that one does not divide by the area S, since the area elements dS have not been taken into account in the summation above. On the other hand, one must divide by the number of points N_p.

2.3.2.2 Taking into account reflections

The free field radiation assumption corresponds to a theoretical case and, in practice, reflections can be expected. Hence, a very simple model of specular reflection has been used.

The interference of the sound wave directly radiated by the open end of waveguide geometry with a wave reflected on a plane perpendicular or parallel to the baffle plane has been implemented. To do so, the expression of the radiated pressure has been modified to add the contribution of N_r reflected waves.

The distance traveled by the reflected wave is represented in Figs. 2.14 for a reflection on a plane respectively perpendicular (Fig. 2.14a) and parallel (Fig. 2.14b) to the baffle. The distance d_r corresponding to the projection of the incident path h_i on the reflection plane can be computed by stating that the incident angle is equal to the reflection angle:

$$d_r = \frac{h_s d}{2h_s - (x_{2r} - x_{2s})}, \tag{2.48}$$

with (x_{1s}, x_{2s}, x_{3s}) the coordinates of the source point, (x_{1r}, x_{2r}, x_{3r}) the coordinates of the reception point and h_s the distance between the source point and the reflection plane. In the case of a reflection plane perpendicular, the distance d corresponding to the projection of the incident and reflected paths h_i and h_r on the reflection plane is $\sqrt{(x_{1r} - x_{1s})^2 + (x_{3r} - x_{3s})^2}$. For a reflection plane parallel, this distance is $\sqrt{(x_{1r} - x_{1s})^2 + (x_{2r} - x_{2s})^2}$.

The incident and reflected distances h_i and h_r can then be computed,

$$h_i = \sqrt{h_s^2 + d_r^2} \ , \tag{2.49}$$

in the case of a perpendicular reflection plane,

$$h_r = \sqrt{(h_s - (x_{2r} - x_{2s}))^2 + (d - d_r)^2} \ , \tag{2.50}$$

and in the case of a parallel reflection plane,

$$h_r = \sqrt{(h_s - (x_{3r} - x_{3s}))^2 + (d - d_r)^2} \ . \tag{2.51}$$

The total path is thus

$$h_t = h_i + h_r \ . \tag{2.52}$$

The contribution of N_r reflected waves can be added to the computation of the radiated pressure at the reception point,

$$p(x_{1r}, x_{2r}, x_{3r}, k) = \sum_{n=0}^{\infty} \frac{j\omega\rho}{2\pi S} \int_S U_n \psi_n(x_{1s}, x_{2s}) \left(\frac{e^{jkh}}{h} + \sum_{r=0}^{N_r} R_r \frac{e^{jkh_{tr}}}{h_{tr}} \right) dS, \tag{2.53}$$

where ω is the angular frequency, ρ is the air density, S is the surface of the open end of the waveguide, (x_{1s}, x_{2s}, x_{3s}) are the coordinates of the points on S, U_n is the modal amplitude of the volume velocity corresponding to the mode ψ_n, $h = \sqrt{(x_{1r} - x_{1s})^2 + (x_{2r} - x_{2s})^2 + (x_{3r} - x_{3s})^2}$ is the distance traveled by the direct wave, h_{tr} are the distances traveled by the reflected waves computed with Eq. 2.52 and R_r are the reflection coefficients.

This integral can be approximated considering a finite summation over $M \times N$ modes and replacing the integral by a finite summation evaluated on a grid of N_p points belonging to the surface S,

$$p(x_{1r}, x_{2r}, x_{3r}, k) = \sum_{n=0}^{N} \frac{j\omega\rho}{2\pi S} \sum_{g=1}^{N_p} U_n \psi_n(x_{1sg}, x_{2sg}) \left(\frac{e^{jkh_g}}{h_g} + \sum_{r=0}^{N_r} R_r \frac{e^{jkh_{trg}}}{h_{trg}} \right) dS, \tag{2.54}$$

with $h_g = \sqrt{(x_{1r} - x_{1sg})^2 + (x_{2r} - x_{2sg})^2 + (x_{3r} - x_{3sg})^2}$ the distance traveled by the direct wave between the point g of the grid and the reception point, h_{trg} the path corresponding to the reflection r of the sound generated by the point g of the grid, $(x_{1sg}, x_{2sg}, x_{3sg})$ being the coordinates of the points of the grid.

2.4 Application to vocal tract

The implementation performed allows one to simulate the acoustic properties of the vocal tract. Straight vocal tract geometries discretized in a concatenation of constant cross-section sections can be simulated from very simple vocal tract models [1], or from area functions [23] extracted from medical imaging procedures.

For every simulation, the impedance or the admittance matrices, the modal amplitude vectors for the pressure and the volume velocity, and the projection matrix are computed at the entrance and the exit of each section. Then, the pressure can be computed at any desired location inside or outside of the vocal tract geometry. This allows one to compute transfer functions and pressure fields. This has been applied in chapters 4 and 5 for the comparison with the experimental data.

Chapter 3

Experimental and numerical methods

Contents

In this chapter, the vocal tract replicas and the experimental setup used to perform experimental measurements of the pressure field and the transfer functions are presented. The signal processing of the experimental data is detailed, and the influence of some experimental artefact is evaluated. Eventually, the FEM simulations also used to compare with the experimental data and the MM simulation data are presented. The data acquired with the experiments and the FEM simulations are compared with data obtained by MM simulations in the chapters 4 and 5.

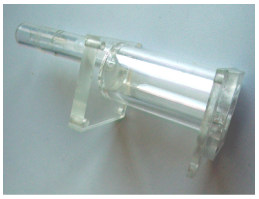
3.1 Vocal tract geometries.

In order to perform the measurements presented in this document, mechanical vocal tract replicas with increasing complexity have been built. Pictures of these replicas are presented in Fig. 3.1. Other geometries have been used for simulations only.

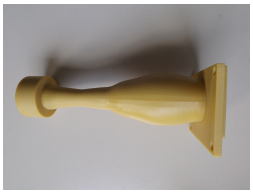
The simplest approximation of the vocal tract has been realized with a uniform tube having a diameter of 29.5 mm and a length of 170 mm. The abscissa $x_3 = 0$ corresponds to the position of the sound source. A baffle is attached to the other end of the replica. It can be

a) One tube

b) Two concentric tubes



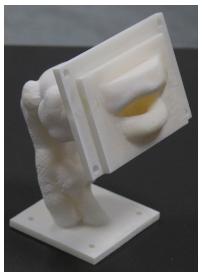
c) Two eccentric tubes



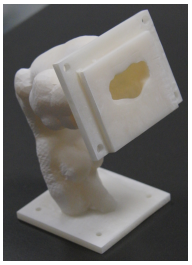
d) 44 concentric tubes

e) 44 eccentric tubes

f) 44 eccentric tubes with elliptical cross-sections



g) MRI with lips



h) MRI without lips

Figure 3.1: Vocal tract replicas. The sound source is connected at the left end of the replicas and at the bottom for the MRI based ones. A baffle is attached to the other end.

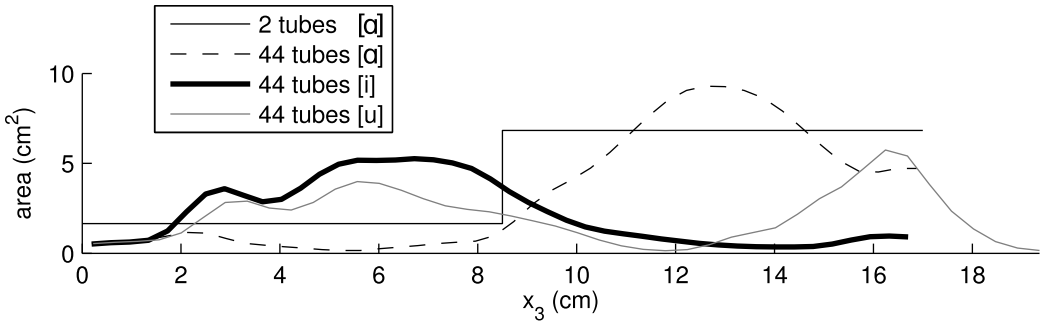


Figure 3.2: Area functions corresponding to the vowels [ɑ], [i] and [u] provided by [1] for the two tubes approximation and by [23] for the 44 tube ones. The abscissa $x_3 = 0$ corresponds to the position of the sound source.

considered as very simple modelisation of the vowel [ə] [1]. A replica of it has been made with a Plexiglass tube (see Fig. 3.1a). It has been used mainly for preliminary tests, see sections 3.3.2, 3.3.3 and 3.3.4.

In order to study the effect of the eccentricity of a junction between two waveguide sections, two geometries have been used.

1. One with two cylindrical tubes sharing the same central axis (see Fig. 3.1b, referred hereafter as the two circular concentric [ɑ], 2CC[ɑ]).
2. One with two cylindrical tubes sharing a common line on the edge (see Fig. 3.1c, referred hereafter as the two circular eccentric [ɑ], 2CE[ɑ]).

The two tubes of both geometries are 85 mm long and have an internal diameter of 14.5 mm and 29.5 mm. The corresponding area function is represented in Fig. 3.2. Although these geometries are a very rough two tube approximation of the vocal tract, they can be seen as a vowel [ɑ] reproduction [1]. The common edge of the second geometry can be seen as a reproduction of the palate. Two replicas of these geometries have been built with Plexiglas tubes in order to perform measurements.

More realistic vocal tract geometries have been designed using the area functions provided by Story [23] for the vowels [ɑ], [i] and [u] (see Fig. 3.2).

Four geometries of increasing complexity have been designed with the area function corresponding to the vowel [ɑ].

1. One has circular cross-sections sharing the same central axis (see Fig. 3.1d, referred hereafter as the 44 cicular concentric [ɑ], 44CC[ɑ]).
2. One has circular cross-sections sharing a common line on the edge (see Fig. 3.1e, referred hereafter as the 44 circular eccentric [ɑ], 44CE[ɑ]).
3. One has elliptical cross-sections sharing a common line on the edge in the plane (x_2,x_3) (see Fig. 3.1f, referred hereafter as the 44 elliptical eccentric [ɑ], 44EE[ɑ]).
4. One has elliptical cross-sections sharing a common line on the edge in the plane (x_2,x_3). The centers of the consecutive cross-sections are shifted by 25% of the maximal possible shift in the plane (x_1,x_3), the edges of the smaller ones being contained inside the larger ones (see Fig. 3.3a). This geometry is referred hereafter as the 44 elliptical fully eccentric [ɑ], 44EFE[ɑ]).

In the case of the 44CE[ɑ], the 44EE[ɑ] and the 44EFE[ɑ], the length of each section has been adjusted so that the total length of the line that connects the centers of each cross-section (referred as the midline hereafter) is the same as for 44CC[ɑ]. Otherwise, numerical tests have shown that a non-expected shift down of the formant positions can be produced (up to about 2%). Thus the length of each section is 3.88 mm, 3.82 mm, 3.86 mm and 3.85 mm for the

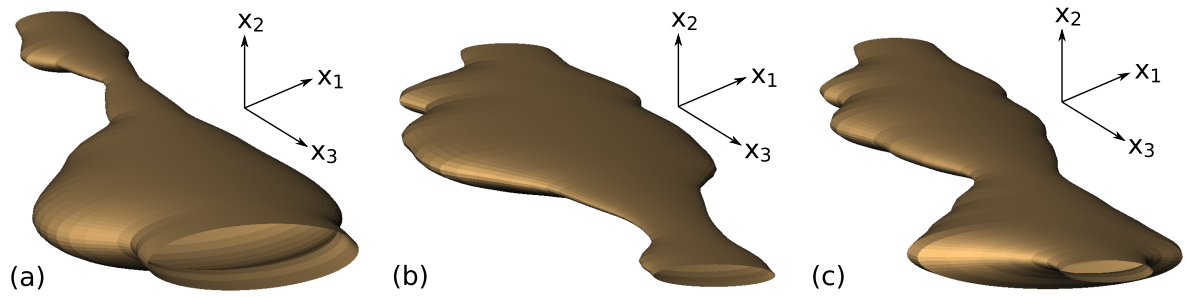


Figure 3.3: Vocal tract geometries used for the simulations (a) vowel [ɑ] (b) vowel [i] (c) vowel [u]

Vowel	[ɑ]	[i]	[u]
Width (cm)	2.21	1.25	0.47
Height (cm)	0.68	0.23	0.11
Ratio	3.25	5.43	4.27

Table 3.1: Width (along x_2), height (along x_1) and ratio of width over height of the elliptical shapes provided by [43] for three vowels.

44CC[ɑ], the 44CE[ɑ], the 44EE[ɑ] and the 44EFE[ɑ], respectively. On the other hand, the eccentricity of the mouth aperture, obtained from [43] for the vowel [ɑ] (see Tab. 3.1), has been applied to all cross-sections to generate the 44EE[ɑ] and 44EFE[ɑ], as done in [18].

Mechanical replicas of the three first geometries corresponding to vowel [ɑ] have been built with a 3D printer (ProJet 3510 SD, with accuracy 0.025 mm to 0.05 mm per 25.4 mm) using a UV curable plastic visijet M3X material (see Figs. 3.1d, 3.1e and 3.1f).

Geometries with elliptical cross-sections sharing a common line in the plane (x_2,x_3) and with 25% of eccentricity in the plane (x_1,x_3) have been created from the vowels [i] and [u] area functions (see Figs. 3.3b and 3.3c). As for the vowel [ɑ], the ratio of the major axis over the minor axis of the elliptical cross-sections have been determined after the work of Fromkin [43] (see Tab. 3.1). The midline of these geometries has also been adjusted so that it is similar to the midline of a geometry without eccentricity. The corrected length of the section is thus 3.81 mm for the 44EFE[i] and 4.34 mm for the 44EFE[u].

In order to have more realistic vocal tract geometries, the MRI-based 3D vocal tract generated by [26] for vowel [ɑ] was adapted for the current work. Some additional elements not necessary for this study were removed, such as the subglottal tube and part of the face. The resulting vocal tract geometries were then set in a rigid flat baffle located at the mouth termination plane, which is defined as the last front-plane that produces a closed outline when it intersects with the vocal tract (see [44]). In order to study the influence of the lips, a geometry without lips has been created. Note that in these geometries the teeth are not present because they were not visible on the MRI.

Mechanical replicas of these geometries were 3-D-printed using the same method as for

the 44 tubes replicas (see Figs. 3.1g and 3.1h).

3.2 Experimental setups

An experimental setup inspired from the work of [19] has been developed to measure the acoustic pressure inside and outside of the vocal tract replicas. Schematic diagrams of this setup are presented in Figs. 3.4a and 3.4c for the planes (x_2, x_3) and (x_1, x_3) respectively. A simplified version dedicated to directivity measurements is presented in Fig. 3.5.

In order to be able to generate sound over the frequency range 100 Hz to 15000 Hz, two different acoustic sources were needed. Because a loudspeaker would not provide enough acoustic energy, compression chambers have become necessary. Two chambers have been used, a Monacor KU-916T and an Eminence PSD:2002S-8 (see Fig. 3.4b) to generate sound in the frequency ranges 100 Hz up to 2000 Hz and 2000 Hz up to 15000 Hz, respectively. The source is connected to the replica with an adaptation part which features a 2 mm diameter centered communication hole from which the sound radiates inside the vocal tract replica (see Fig. 3.4b). Used signals are sines with fixed frequency, linear sweeps, broadband noise and glottal flow simulated with a two mass model [45].

The acoustic pressure is then measured with a B&K 4182 probe microphone equiped with a 1 mm in diameter and either 200 mm or 25 mm long probe. A 3D positioning system (OWIS PS35, see Fig. 3.4d) allows one to move the probe microphone in all directions with an accuracy of $\pm 1.10^{-4}$ m and an amplitude of motion of 250 mm in x_1, 250 mm in x_2 and 400 mm in x_3 (see Fig. 3.4d). However this accuracy can be reduced by the initial positioning which is less accurate (± 1 mm) and can induce a systematic error.

In order to measure the acoustic pressure at greater distances, the positioning system and the probe microphone have been replaced by a microphone (B&K 4192 with a preamplifier B&K 2669 L) placed on a mobile support (see Fig. 3.5b). This allows one to measure the acoustic pressure at 48 cm from the exit of the replica at different angles from the axis of the replica (every 15° from -90° up to 90° as indicated in Fig. 3.5a). The angle 0° corresponds to the center of the exit of the replicas following x_3. For the eccentric geometries the common edge corresponds to the angle 90°.

The microphone signal V_m is next transmitted by means of a microphone conditioner (B&K 5935 L) to a data acquisition card (NI PCI-MIO 16 XE) at a sampling frequency of 44150 Hz. This card can record signals with different ranges of amplitude. The ranges used for the experiments were -5 V to +5 V and -10 V to +10 V, depending on the amplitude of the acoustic pressure and the settings of the other devices. The signals having an amplitude higher than these ranges were saturated. This card is also used to generate the excitation signal which is transmitted to an amplifier (Onkyo a-807) and then to the compression chamber. The input voltage V_s of the sound source is also measured in order to get a phase reference. The whole process is controlled by a Labview program.

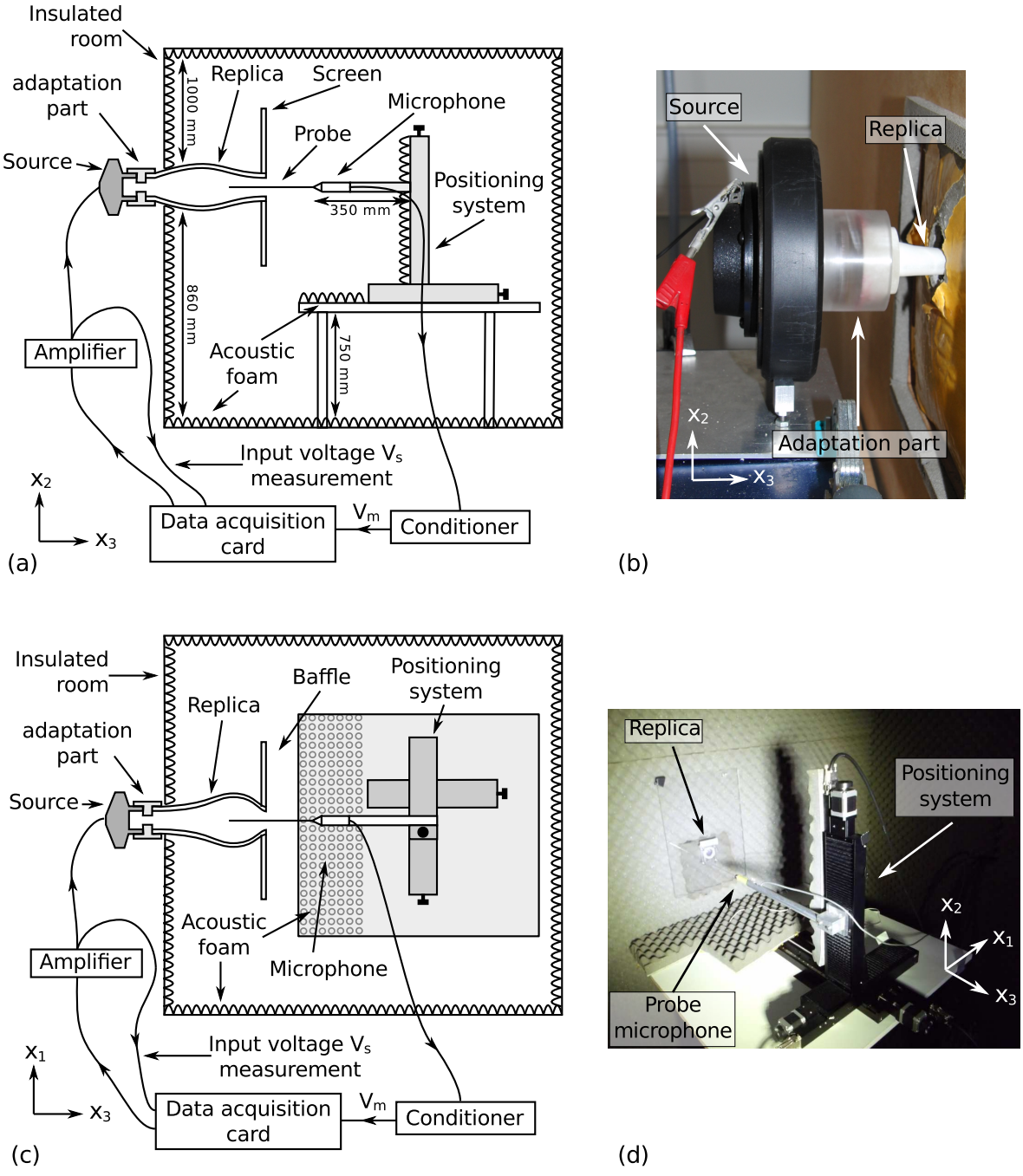


Figure 3.4: Experimental setup (a) schematic diagram in the plane (x_2,x_3) (b) sound source (Eminence PSD:2002S-8 in this case) (c) schematic diagram in the plane (x_1,x_3) (d) positioning system inside the insulated room

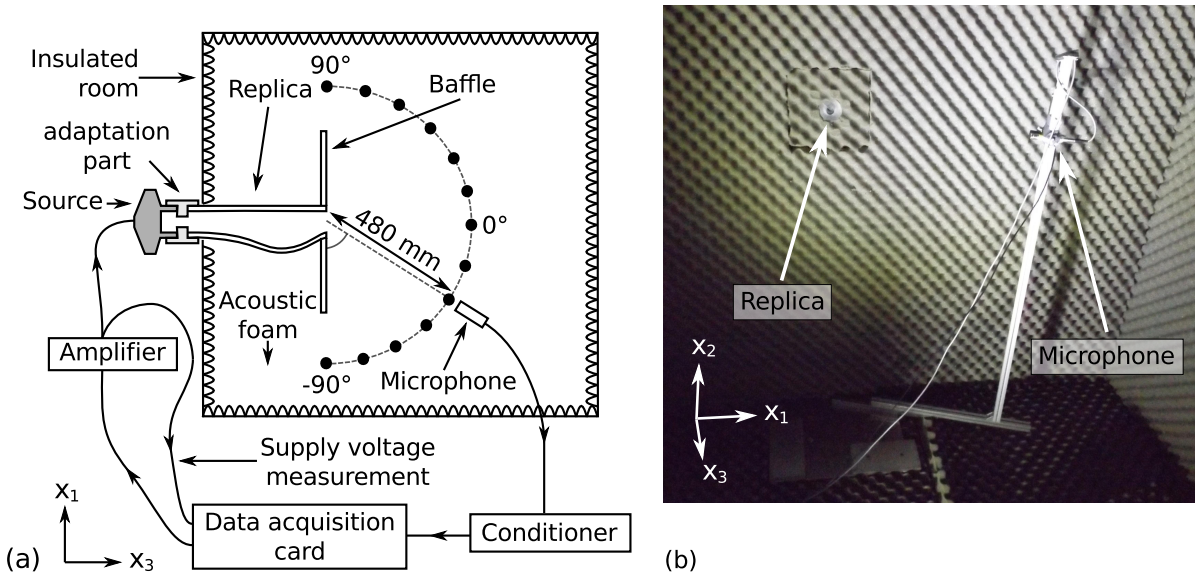


Figure 3.5: Microphone setup, (a) schematic diagram in the plane (x_1,x_3), the eccentricity of the replica corresponds to the experimental configuration (b) picture

The open end of the replica, the microphone and the positioning system are set inside an insulated room [1.92x1.95x1.99 m, Vol = 7.45 m^3, 46]. Even though this room cannot be considered as perfectly anechoic, its acoustic performances in the frequency range of interest in this study (2 kHz to 10 kHz) are sufficient so that the influence of the external noise can be neglected and the free field assumption holds: the direct field is higher than the reverberated field up to 0.94 m from the open end of the replica and the attenuation of the external noise is greater than 25 dB SPL [46].

Acoustic foam is placed on the table on which the positioner is located and on the positioner to minimise reflections. The acoustic source is placed outside of the insulated room to avoid the interference between the sound directly radiated by the compression chamber and the sound radiated by the replica. A rigid rectangular plane screen made of plexiglass is set on the exit of the replica to get large dimensions (365x360 mm) compared to the wavelengths of interest. It corresponds to the flanged boundary condition of the theory.

3.3 Data analysis

In order to measure transfer functions and pressure field maps, sines and linear sweeps have been used as excitation signals. The expression of the signal $s(t)$ corresponding to the linear sweeps is:

$$s(t) = \sin\left(2\pi\left(f_{min}t + \frac{\Delta f}{2\Delta t}t^2\right)\right) \ , \tag{3.1}$$

where f_{min} is the minimal frequency of the sweep signal, Δf is the frequency range swept and Δt is the duration of the signal. Such kind of measurement is performed either at a

limited number of locations and with a broad frequency range for the transfer function measurements or at various locations distributed on a plane with a limited frequency range, or a single frequency, for the pressure field maps measurements. Sinus is also used to measure transfer functions by changing the frequency step by step (typically 10 Hz, each frequency being generated during 1 s). Since these measurements have been used to validate simulation methods, their accuracy and repeatability have been carefully investigated. On the other hand, the reliability of the analysis method and its sensitivity to nonlinear phenomena has been carefully investigated.

3.3.1 Amplitude and phase estimation

The use of sinusoidal signals allows one to concentrate the acoustic energy in one frequency and to perform a measurement as long as desired at this frequency. As a consequence a good signal to noise ratio can be achieved. The use of linear sweep signal with low slope (typically 500 Hz/s) can also ensure a good signal to noise ratio.

However, despite all the care used during the measurements, some phenomena can affect the quality of the recorded signals. One cannot exclude the presence of noise and the harmonic distortion of the sound source cannot be avoided. On the other hand, the signals can feature continuous components and transients can be present when the sound source starts to generate sound and after the frequency changes when the frequency is varied step by step for the sines. Eventually, the physical phenomenon of acoustic streaming [47]–[49], which can be present at the communication hole of the adaptation part or in narrow parts of the replicas, can induce a nonlinear behavior. In order to isolate the effects of these phenomena from the useful part of the signal, a careful signal processing method is used.

The parts of the signals affected by transients are removed. In the case of sines whose frequency is changing step by step, the first 200 ms of each new frequency generation are removed. The biginning and the ending of the linear sweeps are also reduced.

In the case of the linear sweeps, the recorded signal is sliced in small windows (typically 2048 samples) overlapping each other by an overlap rate α determined so that the desired frequency resolution df is achieved:

$$\alpha = 1 - \frac{df \times s_r \times \Delta t}{N_s \times \Delta f} \ ,$$

where s_r is the sample rate, Δt is the duration of the sweep signal, N_s is the number of samples of the window used and Δf the difference between the minimal and the maximal frequency of the signal. A Hann window is applied in order to limit the artefacts due to the small duration of the window [50]. In the case of the sines, the signal is also sliced in windows corresponding to the changes of frequency or position.

The Fourier transform is applied to each window so that one obtains the spectrum of the acquired signal. Zero-padding is used to increase the frequency resolution. The amplitude of the spectrum is normalized multiplying it by $2/N$, N being the number of samples of the

signal. In order to reduce the influence of the background noise and to remove the eventual continuous component and the harmonics generated by the nonlinearities, the maximum of amplitude of the spectrogram is searched inside a reduced frequency interval (100 Hz) centered on the presumed frequency of the signal. Because the point with maximal amplitude does not necessarily corresponds to the maximum, a parabolic interpolation is performed on the three points closest to the maximum in order to estimate the actual maximum and the corresponding frequency. Indeed, small deviations between the desired frequency and the actual one can happen if the sampling frequency of the material is slightly different from its setpoint. A similar interpolation is used to estimate the phase corresponding to the frequency of the maximum. This signal processing method has been validated with synthetic signals.

In order to simplify the computations and the storage of the data, a complex notation $Ae^{j\phi}$ is used for the signals, A being the amplitude and ϕ the phase. The acoustic pressure is estimated from the microphone output signal $V_m(f) = |V_m|e^{j\phi_m}$ and the input voltage of the sound source $V_s(f) = |V_s|e^{j\phi_s}$ as

$$P(f) = \frac{|V_m|}{|V_s|} e^{j(\phi_m - \phi_s)} \ .$$

Given that the sensitivity of the microphone is unknown, and that due to the probe it varies with the frequency, the amplitude of the microphone signal is not converted to Pa. Thus, when it is necessary, the amplitude is normalized by the maximal value, and when it is not possible, the unit is noted $\propto Pa$. This is not disturbing as long as no quantitative evaluation of the data is needed. An additional phase corresponding to the $e^{j\omega t}$ factor can be added to $P(f)$ and thus one as

$$P(f,t) = \frac{|V_m|}{|V_s|} e^{j(\phi_m - \phi_s + \omega t)} \ .$$

Adding this factor allows one to compute the acoustic pressure at any time t instant of the periodic oscillation at the frequency f. This allows one to create animations showing the variations over time of the acoustic pressure at this frequency.

3.3.2 Influence of nonlinearities

In order to evaluate the influence of the nonlinearity of the sound source, induced by the electrodynamic system and or the acoustic streaming, a measurement has been performed varying the amplitude of a sinusoidal excitation signal for a fixed position of the microphone inside the 44CT (at the center of the exit surface).

The ratio of the amplitude $|V_m|$ of the microphone output voltage over the amplitude $|V_s|$ of the sound source input voltage is presented as a function of $|V_s|$ in the Fig. 3.6 for two frequencies. The maximal value has been normalized to 0 dB in each case. One can see that the ratio $\frac{|V_m|}{|V_s|}$ decreases when the input voltage $|V_s|$ increases. It is reduced by 1.5 dB for $|V_s| = 10V$.

This can be explained by the fact that the nonlinearity of the sound source induces harmonic distortion of the signal. When the amplitude is increased the energy of the higher order

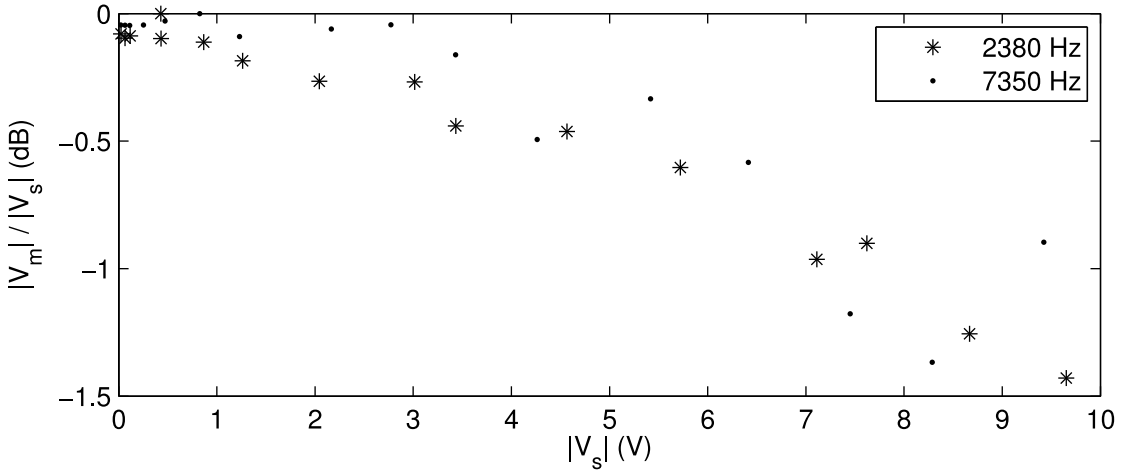


Figure 3.6: Ratio of the amplitude $|V_m|$ of the microphone output voltage over the amplitude $|V_s|$ of the sound source input voltage as a function of $|V_s|$ obtained with two sinusoidal input signals of frequency 2380 Hz and 7350 Hz. The data have been normalized by the maximum.

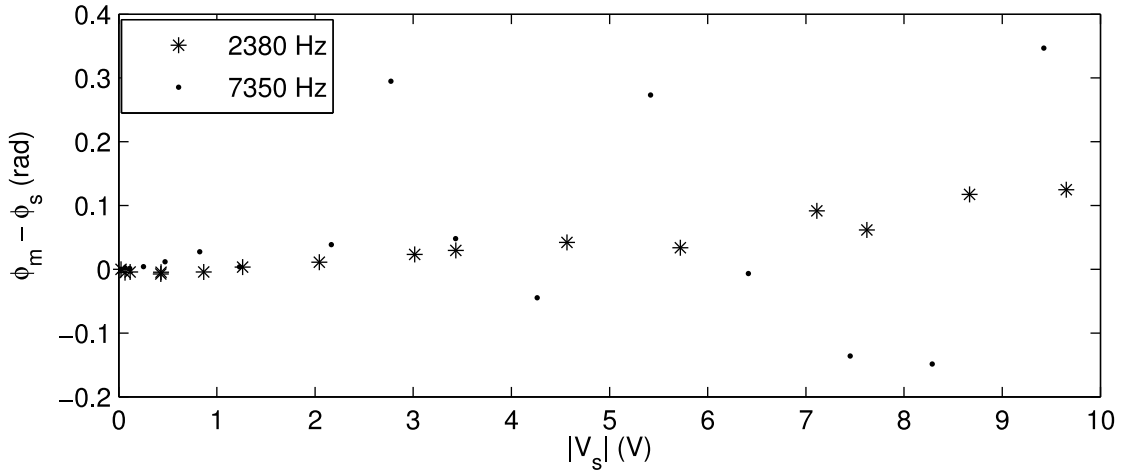


Figure 3.7: Difference between the phase ϕ_m of the microphone output voltage and the phase ϕ_s of the source input voltage as a function of the amplitude $|V_s|$ of the sound source input voltage obtained with two sinusoidal input signals of frequency 2380 Hz and 7350 Hz. The phase corresponding to the lowest $|V_s|$ has been removed to all the data.

harmonics generated by the distortion is increased. Thus, the energy of the first harmonic is decreased and the amplitude of the spectrum at this frequency is decreased. The amplitude $|V_s|$ is not influenced by the nonlinearities and remains the same. Thus, the ratio $\frac{|V_m|}{|V_s|}$ decreases.

The difference between the phase ϕ_m of the microphone output voltage and the phase ϕ_s of the source input voltage is presented as a function of the amplitude $|V_s|$ of the sound source input voltage in Fig. 3.7 for two frequencies. The phase corresponding to the lowest value of $|V_s|$ has been subtracted from all the data.

One can see that the phase difference $\phi_m - \phi_s$ tends to increase when $|V_s|$ is increased. However, it is reduced for four voltage values for the frequency 7350 Hz. The maximal phase deviation is lower than 0.2 rad and 0.4 rad for 2380 Hz and 7350 Hz respectively.

For the input voltage range used (0 - 10 V), the maximal amplitude deviation induced by the nonlinear behavior of the sound source is estimated to be about 1.5 dB. The maximal phase deviation is of the order of 0.4 rad.

The mean value of the input voltage used for the experiments is 3.5 V. This has been determined in order to get a compromise between a sufficient amplitude of the acoustic pressure and a minimal disturbance due to nonlinearities. According to the measurements presented in Figs. 3.6 and 3.7, the maximal disturbance expected from nonlinearities with an input voltage of 3.5 V is about 0.5 dB and 0.4 rad.

The signal processing method (see section 3.3.1) which isolates the first harmonic from the others, allows one to limit the disturbances due to the nonlinearities on the analysis of the data. If, in addition, the amplitude of the input signal of the sound source is kept constant during the measurement, one can use a linear system analysis for the computation of the transfer functions without major disturbance.

3.3.3 Repeatability of measurements

In order to test the repeatability of the measurements performed with both the sines and the linear sweep signals, a measurement has been repeated ten times. The frequency has been varied from 1 kHz to 15 kHz and the acoustic pressure has been recorded at a fixed position at the exit of a one tube replica. The duration of the sweep was 50 s, and for the sines the frequency was increased by steps of 20 Hz.

The ratio $\frac{|V_m|}{|V_s|}$ averaged on the ten measurements is presented in Figs. 3.8a and 3.9a for the sines and the sweeps respectively. The maximal difference between these ratios is presented in Figs. 3.8b and 3.9b. Likewise, the phase difference $\phi_m - \phi_s$ averaged on the ten measurements is presented in Figs. 3.8c and 3.9c for the sines and the sweeps respectively, and the maximal difference between the different phase measurements is presented in Figs. 3.8d and 3.9d.

The variations of the amplitude ratio with the frequency can be attributed to the combi-

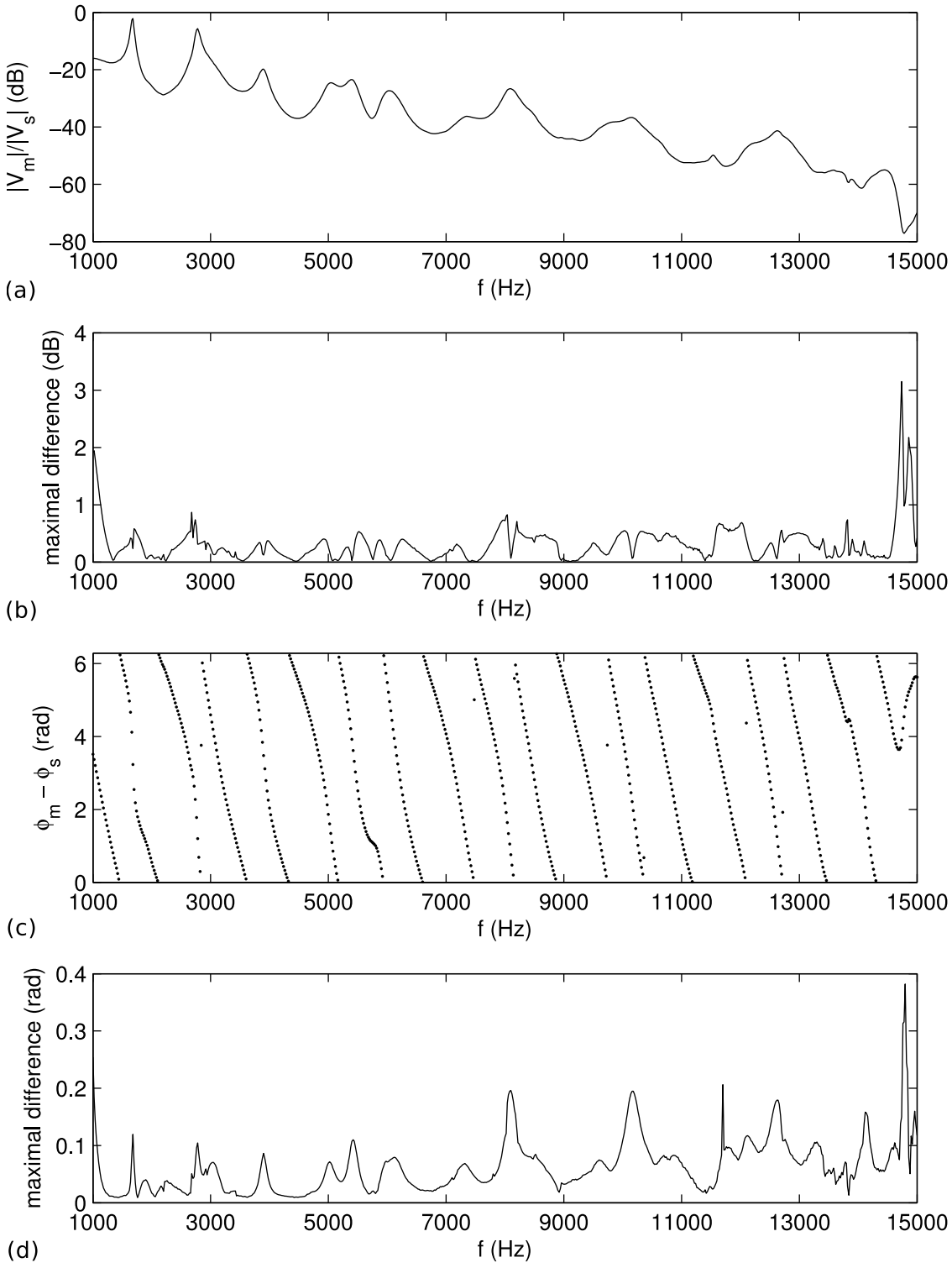


Figure 3.8: (a) Ratio of the amplitude of the microphone signal $|V_m|$ over the sound source input voltage $|V_s|$ averaged on 10 measurements performed at the same location with a sines signal whose frequency is varying by steps 20 Hz from 1 kHz to 15 kHz (b) maximal ratio difference between the 10 measurements (c) phase difference $\phi_m - \phi_s$ between the microphone signal and the sound source input signal averaged on the 10 measurements (d) maximal variation of the phase difference between the 10 measurements.

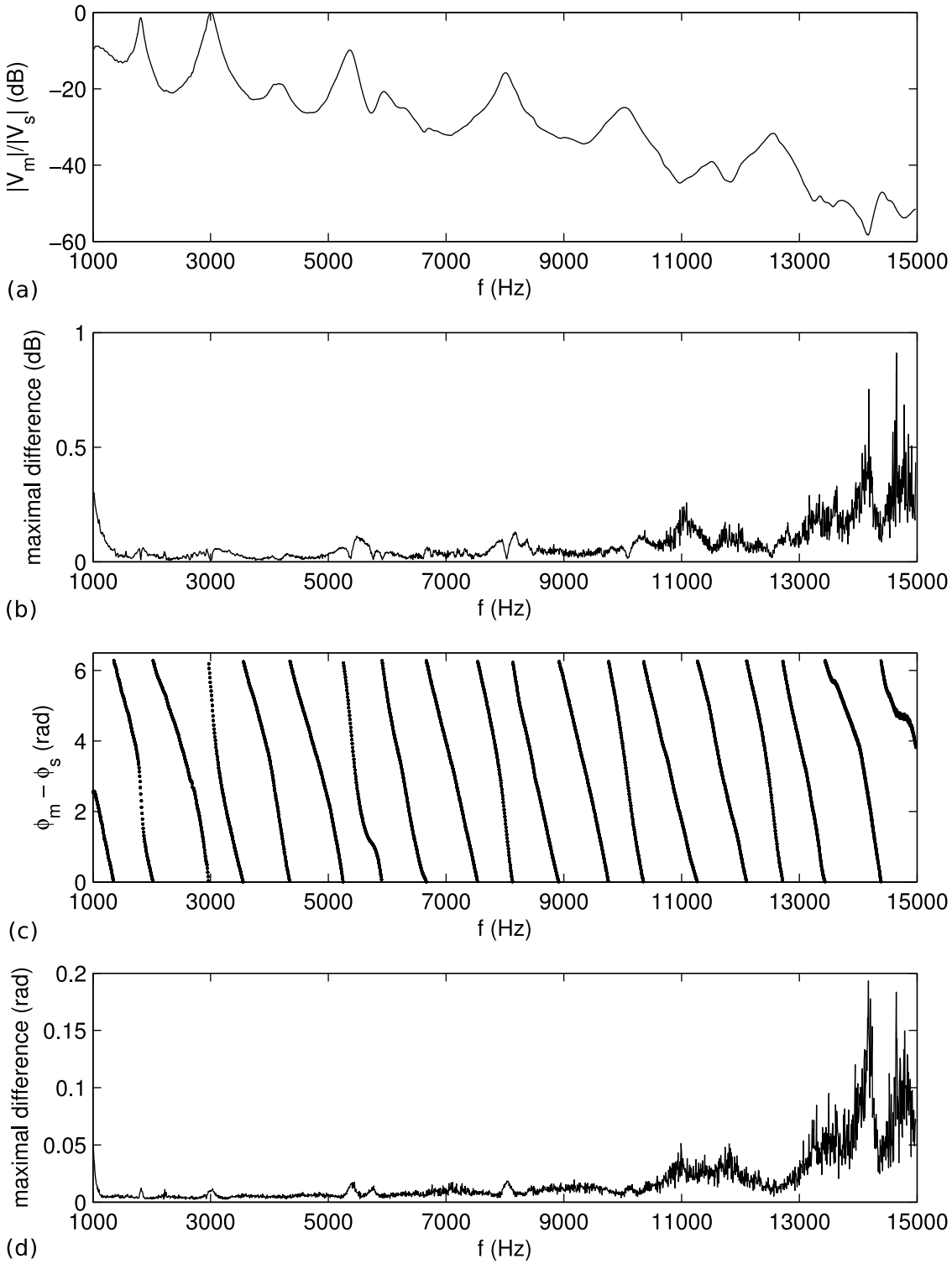


Figure 3.9: (a) Ratio of the amplitude of the microphone signal $|V_m|$ over the sound source input voltage $|V_s|$ averaged on 10 measurements performed at the same location with a sweep signal whose frequency varies from 1 kHz to 15 kHz in 50 s (b) maximal ratio difference between the 10 measurements (c) phase difference $\phi_m - \phi_s$ between the microphone signal and the sound source input signal averaged on the 10 measurements (d) maximal variation of the phase difference between the 10 measurements.

nation of the effect of the acoustic resonances of the tube and the frequency dependence of the sound source amplitude. For the sines, the maximal amplitude difference is smaller than 0.9 dB in the range 1.2 kHz to 14.5 kHz. Outside this range it can reach up to 3 dB. The maximal phase difference is smaller than 0.2 rad in the range 1 kHz to 14.5 kHz, and beyond 14.5 kHz it goes up to 0.39 rad. For the sweeps, the maximal amplitude difference is smaller than 0.26 dB in the range 1.2 kHz to 13 kHz, and it can reach 1 dB out of this range. The maximal phase difference is smaller than 0.06 rad in the range 1 kHz to 13 kHz, and beyond 13 kHz it can reach 0.2 rad.

The phase variations of the sines and the amplitude and phase variations of the sweeps globally increases with the frequency. This can be understood as the consequence of the global decrease of the amplitude of the measured acoustic pressure when the frequency is increased. Indeed, the sound source becomes less efficient at high frequency, and the probe of the microphone by damping the high frequencies acts as a low pass filter. The lower signal to noise ratio at high frequency induces more variations between the different measurements. However, the amplitude variations of the sines remain globally constant in the frequency range 1.2 kHz to 14.5 kHz.

The greater variations observed between 1 kHz and 1.2 kHz can be related to the fact that the optimal frequency range recommended by the constructor of the sound source Eminence PSD:2002S-8 is 1.2 kHz to 20 kHz. Thus, the sound source may not provide enough acoustic energy between 1 kHz and 1.2 kHz to achieve a signal to noise ratio as high as between 1.2 kHz and 13 kHz.

The fact that greater variations (0.9 dB and 0.2 rad) are observed with the sines than with the sweeps (0.26 dB and 0.06 rad) can appear as unexpected. Indeed, the fact that more acoustic energy is provided at a fixed frequency during a longer time than with the sweep signal is expected to ensure a better signal to noise ratio. However, this can be explained by the fact that the duration of the measurement with the sines is longer than the one with sweeps: the first one lasts about two hours and the second one about 10 minutes. It is hypothesized that the experimental conditions, and more particularly the temperature, can more significantly change during the experiment of two hours than during the one of ten minutes. As a consequence the frequency and the amplitude of the peaks of the measured acoustic pressure can have more variations. Thus, in the case of the sines, the variations of the measured acoustic pressure would not only be due to measurement noise but also to actual variation of the measured quantity. This hypothesis is reinforced by the fact that variations of the maximal difference with the frequency show smooth increases and decreases which could be attributed to frequency shifts of the peaks of the curve rather than to measurement noise. On the opposite, in the case of the sweeps, the variations of the maximal difference with the frequency is more noisy and could be attributed to the measurement noise mainly. If the formulated hypothesis is correct, this show that the duration of a measurement process can impacts its accuracy.

In addition, the amplitude of the maxima of the sines continuously changes along the experiment, and the frequency continuously increases which is consistant with an increase of temperature. It would be possible to extract the temperature curve from the experiment.

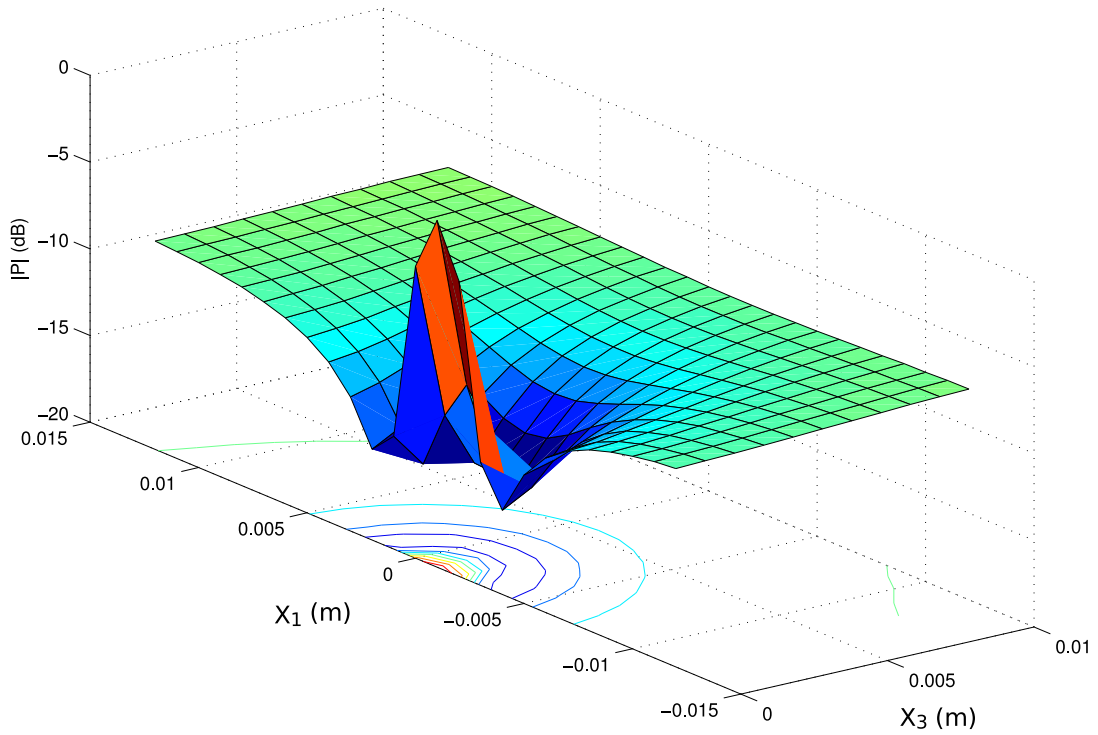


Figure 3.10: Normalized amplitude of the pressure field measured with a spacing of 1 mm on a 20 mm × 10 mm surface perpendicular to the tube axis just in front of the communication hole inside the one tube replica at a frequency of 2550 Hz.

A similar experiment performed on the frequency range 2 kHz to 10 kHz, which confirms these results, is presented in appendix A.

In summary: The variations of amplitude and phase are frequency dependent and they are lower than 0.9 dB and 0.2 rad for the frequency range 1.2 kHz to 14.5 kHz. Thus, the repeatability of the measurements performed with sinusoidal and sweep signals can be considered as good.

3.3.4 Measurements close to discontinuities

The first attempts to measure transfer functions were performed using two points, one located near the exit of the replicas and one near the communication hole of the adaptation part. Even though this was performed under the first cut-on frequency of the replica, discrepancies between the plane wave theory and the experiments were observed. Differences in the frequency of the maxima greater than 100 Hz were observed. In order to understand better the origin of these deviations, the acoustic field has been measured near the communication hole.

The Fig. 3.10 shows the normalized amplitude of the acoustic pressure measured on a 20

mm × 10 mm rectangular surface in the plane (x_1,x_3) just in front of the communication hole inside the one tube replica at a frequency of 2550 Hz. One can see a peak of amplitude close to the exit of the communication hole (at $x_1 = -1$ mm and $x_3 = 0$ mm) surrounded by a minimum. In this small area, variations up to 10 dB within a distance of 1 mm are observed. Close to the communication hole, the plane wave assumption does not hold: variations of amplitude of the order of 15 dB are observed on x_1 perpendicularly to the propagation axis x_3. However, these variations along x_1 decreases when the distance from the communication hole is increased. At $x_3 = 10$ mm, the variations along x_1 are lower than 1 dB and it is possible to consider that the plane wave assumption is valid.

The communication hole being small (2 mm in diameter) compared to the diameter of the replica (29.5 mm), the observed pressure field can be interpreted as the radiation from the communication hole inside the replica. At the frequency of the experiment, 2550 Hz, no HOM can propagate inside the replica (the first cut-on frequency is 6.8 kHz at 26.5 °C). Thus, the observed pressure field results from the excitation of evanescent HOM at the discontinuity. This is in agreement with the fact that the variation along x_1 are almost not noticeable after $x_3 = 10$ mm.

One can notice that the maximum of the radiation pattern is shifted by -1 mm on x_1 from the center of the landmark. Thus, the actual position of the communication hole is at x_1 = -1 mm on this landmark. This can be attributed to an error in the positioning of the probe, and it shows that the positioning error can be of the order of 1 mm.

In summary: the pressure field can have important transverse variations (up to 15 dB) near discontinuities, even at low frequency and the plane wave assumption does not hold inside a small region around the discontinuity. The important variations within a small space interval (up to 10 dB within 1 mm) can increase the effect of the positioning error. The discrepancies observed between the measured and simulated transfer functions came from the fact that the plane wave assumption does not hold at one of the measurement points and that the measurement error is increased when the pressure field has great variations within short space intervals. Measurements close to strong discontinuities, such as the transition from the adaptation part to the replica, should be avoided in order to limit the measurement errors.

3.3.5 Pressure-pressure transfer function

In order to compare the simulations with the experiments and to study the frequency behavior of the vocal tract replicas, pressure-pressure transfer functions (PPTF) have been measured. This kind of transfer function corresponds to the ratio of the acoustic pressure at two different points located inside or outside the replicas. In contrast to a volume velocity-pressure transfer function (VPTF), PPTF do not give direct information about the interaction between a volume velocity sound source, such as the vocal folds, and the vocal tract. Even though this would have been of particular interest in the context of the study of vocal tract properties, it has been chosen to measure PPTF because VPTF is difficult to measure accurately: a direct measure of volume velocity is difficult to implement and with low accuracy, and indirect

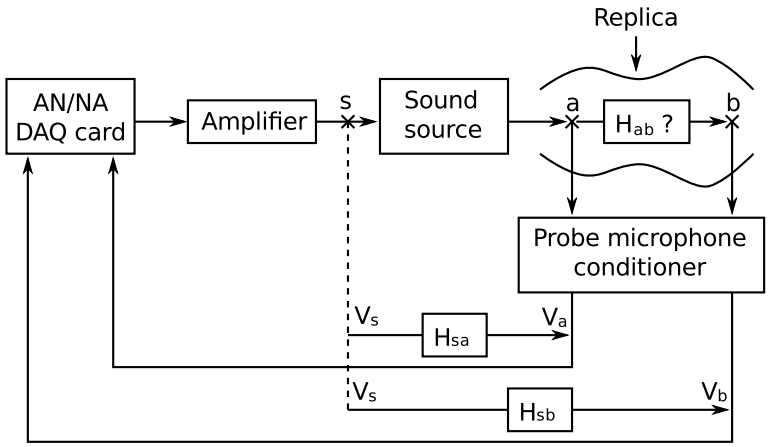


Figure 3.11: Schematic diagram of the transfer function measurement method.

measurement of volume velocity is subject to uncertainties which reduces the accuracy and can introduce experimental bias. In addition, the inaccuracy of volume velocity measurement can be increased when it is measured at the entrance of the vocal tract, as it is the case if one desires to investigate the interaction between vocal folds and vocal tract, because the strong discontinuity present there induces large variations of the acoustic field which can potentially introduce errors linked to positioning inaccuracy (see section 3.3.4).

In contrast, the PPTF rely only on direct measurement of the acoustic pressure with a microphone which is reliable and provides a good accuracy. In addition, as one does not focus on coupling between vocal folds and vocal tract, measuring pressure at the entrance of the vocal tract is not meaningful and one can measure it avoiding strong discontinuities.

In order to measure a PPTF between two points a and b inside or outside a vocal tract replica, the acoustic pressure is measured either using sinusoidal signals varying the frequency step by step or using sweep signals. A schematic diagram of the transfer functions involved in the measurement is presented in Fig. 3.11.

Both pressure measurements are performed with the same probe microphone at two different stages separated in time. This allows one to avoid the calibration of the probe microphone which can induce additional uncertainties. The microphone output signal V_a is first measured on the whole frequency range studied at the first point of coordinate (x_{1a}, x_{2a}, x_{3a}), then V_b is measured at the second point of coordinate (x_{1b}, x_{2b}, x_{3b}). During each measurement the supply voltage V_s is measured simultaneously at the input of the sound source in order to have a phase and amplitude reference. The acquired signals can be written in complex notation:

$$\begin{cases} V_s = A_s e^{j\phi_s} \\ V_a = A_a e^{j\phi_a} \\ V_b = A_b e^{j\phi_b} \end{cases} . \tag{3.2}$$

Both transfer functions H_{sa} and H_{sb} between the supply voltage V_s and the probe micro-

phone output signals V_a and V_b corresponding to both measurement points are then estimated. To achieve this the amplitudes $|V_a|$ and $|V_b|$ of the signals measured by the probe microphone is divided by the supply voltage amplitude $|V_s|$ to compute the modulus. The phase is obtained by computing the phase shifts $\phi_a - \phi_s$ and $\phi_b - \phi_s$ between the signals measured by the probe microphone and the supply voltage. So the transfer functions H_{sa} and H_{sb} can be expressed in the following way:

$$\begin{cases} H_{sa} = \frac{|V_a|}{|V_s|} e^{j(\phi_a - \phi_s)} \\ H_{sb} = \frac{|V_b|}{|V_s|} e^{j(\phi_b - \phi_s)} \end{cases} \tag{3.3}$$

The transfer function H_{ab} between the measurement points a and b is obtained as the ratio H_{sb}/H_{sa}. The transfer function H_{sa} corresponds to the product of the transfer functions of the sound source, the propagation of sound from the sound source to the point (x_{1a}, x_{2a}, x_{3a}), the probe microphone and the microphone conditioner (see Fig. 3.11). Likewise, H_{sb} corresponds to the product of the transfer functions of the sound source, the propagation of sound from the sound source to the point (x_{1b}, x_{2b}, x_{3b}), the probe microphone and the microphone conditioner. If the experimental conditions are exactly the same (constant temperature) for the measurement of transfer function H_{sa} and H_{sb}, the transfer function H_{sb} is the product of transfer function H_{sa} by the transfer function H_{ab} which ones wants to estimate. Thus we have:

$$H_{ab} = \frac{H_{sb}}{H_{sa}}. \tag{3.4}$$

The transfer functions corresponding to the sound source, the acoustic propagation from the source to the point a, the probe microphone and the microphone conditioner being present in both H_{sb} and H_{sa}, they are eliminated in Eq. (3.4) and their influence on the estimation of H_{ab} is thus limited. This method relies on the hypothesis that the whole measurement system is linear. As mentioned in part 3.3.2, this is not the case for the sound source. However using the signal processing method detailed in 3.3.1 and keeping constant the amplitude of the input voltage of the sound source during the whole process allows one to limit the disturbances due to nonlinearities.

It is to be noted that the peaks that could appear in a PPTF do not correspond to the vocal tract resonances or formants typically observed in vocal tract transfer functions (see, e.g., [51], for an explanation of mechanical transmissibility functions). For instance, if P_a and P_b are, respectively, obtained at the mouth exit and within the vocal tract, P_a will contain all the formants that appear in a vocal tract transfer function, while some of them will also be present in P_b, depending on their pressure distribution. Therefore, note that shared resonances will be canceled once H_{ab} is computed. Thus, the PPTF presented hereafter are discussed in term of peaks and dips and not formants. However, this magnitude is of special interest for the comparison between simulations and experiments, since the microphone and source calibration can be avoided, a single microphone can be used for the measurements and the

influence of the experimental setup (amplifier, wires, and microphone conditioner) on them can be neglected.

3.3.6 Pressure field maps

It is interesting to compare the acoustic pressure measured and simulated on a grid of points distributed on a plane (typically the points of the grid were spaced by 2.5 mm). Even though it is possible to generate an animation of the evolution of the pressure field with the frequency, one can create an image of the pressure field maps only at a fixed frequency. Moreover, similar pressure field patterns are obtained for similar particular phenomenon occurring inside the vocal tract geometries (resonance or anti-resonance) which do not necessarily occur at exactly the same frequency in the simulation and the experiment. It is thus more meaningful to select one of these particular phenomena, which can appear as a peak or a dip of pressure amplitude at a given frequency and to compare the pressure maps at the frequency corresponding to this phenomenon in the experiment and the simulation respectively. That is why different frequencies are used for the experiments and the different simulation methods in the comparisons provided in part 4. More practically, the procedure used for these comparison was to select a peak on the acoustic pressure recorded at one point inside or outside the replica and to extract the pressure field map at the corresponding frequency. The acoustic pressure was then simulated at the same location and the pressure field map was extracted at the frequency corresponding to the same peak of acoustic pressure. These frequencies are thus different from the PPTF peak frequencies. When one considers a peak of amplitude, the perturbations due to the frequency dependence of the sound source amplitude and the various phenomena which can perturb the measurement do not affect the frequency because enough acoustic energy is provided at this frequency.

3.3.7 Modal projection

In order to extract the modal amplitude $\boldsymbol{P}$ from the experimental data, a modal projection has been applied. Multiplying both sides of Eq. (2.4) with $\boldsymbol{\psi}^\dagger$ and integrating over the surface allows one to get

$$\int_S \boldsymbol{\psi}^\dagger p \; dS = \int_S \boldsymbol{\psi}^\dagger \boldsymbol{\psi}^t \boldsymbol{P} dS \; . \tag{3.5}$$

The modal amplitude of the pressure $\boldsymbol{P}$ is independent of the position on the cross-section surface and can be removed from the integral. The normalisation condition (2.5) can then be applied and the relation

$$\boldsymbol{P} = \int_S \boldsymbol{\psi}^* p \; dS \tag{3.6}$$

can be derived. The modal amplitudes $\boldsymbol{P}$ can thus be extracted from the pressure field p.

In practice, the pressure field p_i is measured on a distribution of N discrete points with coordinates (x_{1i}, x_{2i}, x_3) on the cross-section surface. The measurement plane being perpendicular to the axis of the replica, x_3 remains constant. Eq. (3.6) is approximated by a finite

summation

$$P_n(x_3) = \sum_{i=1}^{N} \psi_n(x_{1i}, x_{2i}) p(x_{1i}, x_{2i}, x_3) \tag{3.7}$$

It is assumed that the propagation inside the replicas with circular cross-sections can be described with Eq. (2.6). In this case, Eq. (3.7) can be expressed as

$$P_n(x_3) = \sum_{i=1}^{N} J_n(r_i \gamma_{mn}/R) e^{jn\theta_i} p(r_i, \theta_i, x_3) \ , \tag{3.8}$$

with r_i and θ_i the polar coordinates of the points. This method is applied in section 4.1.4.

3.3.8 Directivity analysis

Directivity measurements are performed by measuring the acoustic pressure at various locations distributed on an half circle whose center is located at the exit of the studied replica. This has been done firstly with the experimental setup of Fig. 3.5 in order to be able to measure at greater distances. Afterwards, the setup of Fig. 3.4 has also been used in order to increase the angular resolution because some observed phenomenon showed significant variations within small angular regions. Directivity maps can be created from these measurements by representing the amplitude of the acoustic pressure as a function of the frequency and the angular position in color scale.

In order to enhance the change of pressure amplitude with respect to the angular position, unless stated differently, the amplitude at the different positions has been normalized by the amplitude at the central positions (which corresponds to the point located in front of the replica at an angle of 0°). The fact that the sound source amplitude depends on the frequency could disturb the interpretation of the measurements and their comparison with the simulations. However, its amplitude being the same for all the positions, dividing by the amplitude at one position cancels its influence, and the results can be compared with the simulations. Thus, one obtains a normalized directivity map.

To get a more general view of the directivity effects, the maximal sound pressure level difference (MSPLD) between the different positions has been computed for all the frequencies. To do so, for each frequency, the minimum of the sound pressure level (in dB SPL) has been subtracted from the maximum with respect to the angular position.

On the experimental data measured with the setup presented in Fig. 3.5 the amplitude at -90° and 90° is lower for every measured frequency. Since it is not the case for the simulated data and that this happens even at low frequencies, this can be attributed to a measurement artifact probably related to the fact that the microphone was close to the wall of the insulated room at these positions. To avoid the perturbations of this artifact in the analysis, the MSPLD has been plotted with and without the measurements of the edges (at -90° and 90°). However the observed overall tendency remains similar.

3.4 FEM

In the framework of the EUNISON project, the measurements and MM simulations performed have been compared with finite element (FEM) simulations performed at the university La Salle of Barcelona.

The finite element method (FEM) was used to solve the time domain wave equation for the acoustic pressure,

$$(\partial_{tt}^2 - c^2\nabla^2)p = 0 \ , \tag{3.9}$$

where $p(x,t)$ stands for the acoustic pressure, c for the speed of sound, and ∂_{tt} for the second order time derivative [18]. First, each one of the vocal tract geometries were set in a rigid baffle with dimensions 0.3 m $\times$ 0.3 m. This baffle constitutes one side of a rectangular volume of 0.3 m $\times$ 0.3 m $\times$ 0.2 m in size which allowed sound waves to radiate out from the vocal tract. Free-field radiation conditions were then emulated by means of a perfectly matched layer (PML) of width 0.1 m that surrounded this radiation space and absorbed the outgoing sound waves. The PML was configured to get a reflection coefficient of $r_\infty = 10^{-4}$ at its outer boundary. The resulting computational domains were then meshed using linear tetrahedral elements with a size that ranged from 0.002 mor 0.001 m within the vocal tract and immediate outer space, to 0.0025 m or 0.005 m in the radiation space and 0.0075 m in the PML region. As far as boundary conditions are concerned, a Gaussian pulse of the type [18]

$$GP(n) = e^{[(\Delta t\ n - T_{GP})0.29T_{GP}]^2}[m^3/s] \ , \tag{3.10}$$

with $T_{GP} = 0.646/f_0$ and $f_0 = 10$ kHz, was imposed at the vocal tract entrance to compute transfer functions, while a sinusoidal signal, i.e., $\sin(2\pi ft)$, with f corresponding to the frequency of interest was used for obtaining pressure maps (see sections 4.1.4 and 4.2.4). To avoid numerical errors beyond the maximum frequency of interest ($f_{max} = 10$ kHz), the Gaussian pulse was filtered using a low-pass filter with cutoff frequency 10 kHz. At the vocal tract walls, a constant frequency boundary admittance coefficient of $\mu = 0.0025$ or $\mu = 0.005$ was imposed to introduce losses. The first value was deemed appropriate after a tuning process, where the resonance bandwidths obtained from simulations were adjusted to fit those from experiments. The second value corresponds to the impedance of the vocal tract tissue $Z_w = 83\ 666$ kg/m^{2}s (see [52]). The baffle where the vocal tracts were set was considered rigid, i.e., $\mu = 0$ on its surface. A finite element simulation lasting 20 ms or 25 ms for each case was then performed with a sampling rate of $f_s = 1/\Delta t = 2000$ kHz or $f_s = 1/\Delta t = 8000$ kHz. Such a high frequency rate is needed to fulfill within the meshed computational domains, a restrictive stability condition of the Courant–Friedrich–Levy type. The value of the sound speed was chosen accordingly to the temperatures of the experiments. It was ranged between 340 m/s and 350 m/s. Simulation times ranged between 24 and 80 hours in a serial computing system with processor Intel® Core™ i5 2.8 GHz. Details on the finite element formulation and on the derivation of the boundary conditions can be found in [18] and [53], respectively.

Chapter 4

Influence of HOM on pressure field and transfer functions

In this chapter, the influence of HOM on the vocal tract acoustics is investigated through PPTF and pressure maps measurements as well as simulated VPTF and input impedance. Using the MM presented in chapter 2 and the experimental and numerical methods detailed in chapter 3, different geometrical approximations of the vocal tract shape corresponding to the vowel [ɑ] presented in section 3.1 are studied and compared. The results presented in this chapter are partly published in [20] and [54].

4.1 Two cylindrical tubes approximation

The two tubes geometries (2CC[ɑ] and 2CE[ɑ]) are very simple approximations of the vowel [ɑ] proposed by Fant [1] in which the excitation and propagation of HOM can easily be understood. That is why it has been extensively studied with both experiments, MM and FEM simulations.

4.1.1 Propagation modes inside the geometries

Inside the small tube of the two tubes geometries the cut-on frequency of the first HOM (the only one propagating under 20 kHz) is 14.5 kHz. The discontinuity corresponding to the transition from the communication hole to the replica is likely to generate evanescent HOM at the entrance of this tube. However, its length being more than five times larger than its diameter, the influence of evanescent HOM can be neglected at the junction between the small tube and the larger one. Thus, between 0 kHz and 10 kHz, only plane waves (mode ψ_{00}) arrive at this junction. Inside the larger tube, ψ_{01} (see Fig. 2.2) which has a cut-on frequency of 6.8 kHz, is the only HOM which can propagate under 10 kHz. Thus, the only modes propagating inside these geometries under 10 kHz are ψ_{00} and ψ_{01}.

The exit of the geometries being perfectly axisymmetric, the HOM ψ_{01} can be excited only if the junction is asymmetric. Indeed, in Eq. (2.36) the contributions of ψ_{00} on either side of the nodal line of ψ_{01} compensate each other exactly if the junction is perfectly axisymmetric. In this case, the coupling term $F_{00,01}$ between ψ_{00} and ψ_{01} is zero and ψ_{01} cannot be excited.

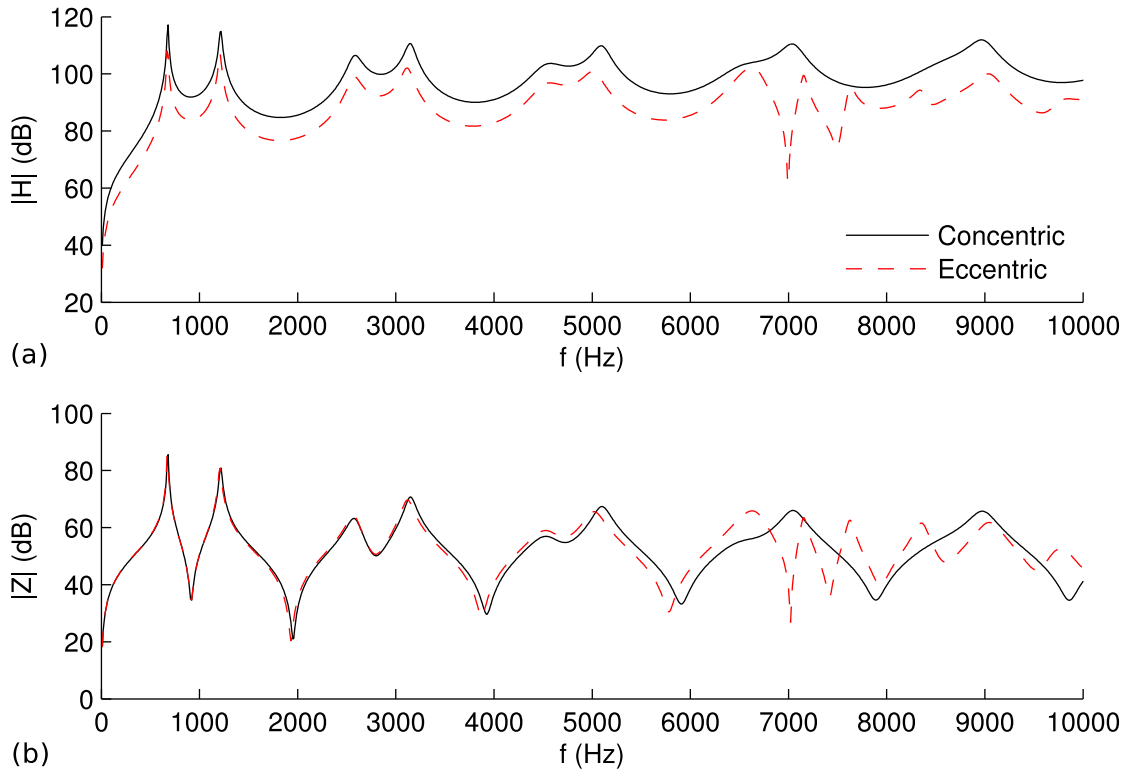


Figure 4.1: Amplitude of the input impedance $|Z|$ and of the transfer function $|H|$ between the input volume velocity and the acoustic pressure radiated at 60 cm in front of the exit of two vocal tract approximations for the vowel [ɑ] based on [1] with a concentric and an eccentric junction between both sections.

If an asymmetry is introduced, as shown in section 2.2.4, $F_{00,01}$ becomes non zero and ψ_{01} can be excited. When the eccentricity is increased $F_{00,01}$ increases. Thus ψ_{01} is not expected to be present in the axisymmetric configuration (2CC[ɑ]).

4.1.2 Simulated velocity-pressure transfer functions and input impedance

In order to characterize the two tube geometries with quantities relevant for speech studies, the input impedance Z and the transfer function H between the volume velocity imposed at the entrance and the pressure radiated at a distance of 60 cm in front of the geometry (angle of 0°), which is relevant in the context of a conversation, have been computed with the MM. The input impedance Z has been computed as the ratio of the acoustic pressure over the particle velocity at the center of the communication hole. This quantity can be useful to characterize the interaction between the vocal folds and the vocal tract while the transfer function helps understanding what a listener can hear with a particular vocal tract geometry. The amplitudes of H and Z are presented for the two types of junctions in Fig. 4.1.

First, different resolutions of the grid used for the MM simulations have been tested by computing the transfer function of the 2CE[ɑ]. The averaged amplitude and phase differ-

Compared resolutions (mm)	2 - 1	1 - 0.5	0.5 - 0.3
Averaged amplitude difference (dB)	3.6	1.9	0.7
Averaged phase difference (rad)	0.8	0.4	0.2

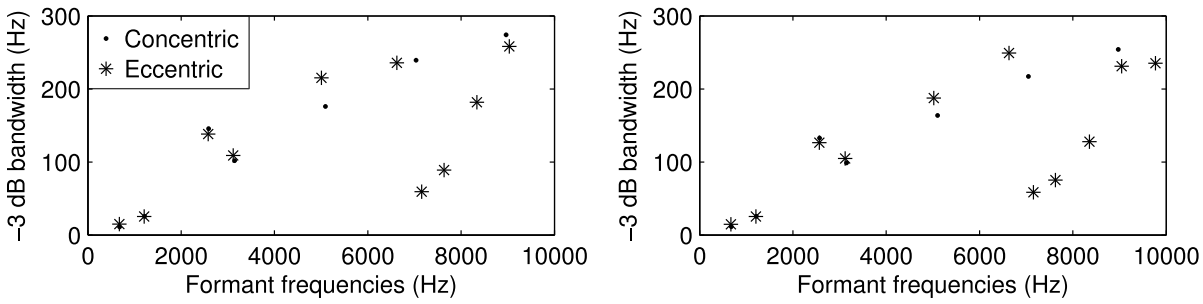
Table 4.1: Averaged amplitude and phase differences between the transfer functions computed with different grid resolutions on a two tube geometrical approximation of the vowel [ɑ] with eccentric junction and circular cross-sections (2CE[ɑ]).

ences between the transfer functions computed with the different resolutions are presented in Tab. 4.1. One can see that the averaged difference reduces when the resolution is increased. Thus, the transfer function obtained with the MM implementation converges. However, it is to be noted that when more resolution steps are tested, the transfer function does not monotonically converge. This can be explained by the fact that the accuracy of the description of the contour of the cross-section does not increase monotonically when the resolution is increased. The resolution 0.5 mm has been chosen for the simulations as a compromise between accuracy and computation time.

The peaks of H and Z corresponds to resonances inside the vocal tract, usually referred to as formant in speech study. As expected, the frequencies of the peaks of Z corresponds to the peaks of H. However, there is a difference of almost 50 Hz (1.2%) for the fifth peak for the 2CC[ɑ], and of almost 80 Hz (1.4%) for the last peak of 2CE[ɑ]. The averaged difference between the frequencies of the peaks of Z and H is 12 Hz for 2CC[ɑ] and of 16 Hz for the 2CE[ɑ]. The anti-resonances observed for the 2CE[ɑ] correspond to dips of Z. The two first formants, which characterize the vowel, have a frequency of 670 Hz and 1200 Hz. These values do not vary on the junction type.

Except for the two last formants of 2CC[ɑ], the frequencies of the formants are lower for the 2CE[ɑ] for both Z and H (down to 1.7% for the sixth formant). This shift down in the formant frequencies could be explained by the fact that the 2CE[ɑ] behaves as if it is extended by the eccentricity of the junction. Indeed the midline is longer for the 2CE[ɑ]. The fact that the two last formants of the 2CC[ɑ] have lower frequencies than the corresponding ones in the 2CE[ɑ] (up to 1.69 % for the seventh formant) can be related to the fact that their frequency is above the cut-on frequency of ψ_{01} (6.897 kHz at 26.5°C) which is expected to be excited in the 2CE[ɑ].

Up to 6.5 kHz the input impedance curves of the concentric and eccentric configurations are almost exactly similar. Above this frequency, the variations of Z are more complex for the eccentric configurations. On the contrary, the transfer function of 2CC[ɑ] has globally an higher amplitude than the one of 2CE[ɑ]: the averaged difference between both transfer functions in the frequency range 0-10 kHz is 9 dB. This could be due to differences in the interaction between the plane wave coming from the small tube and the evanescent HOM excited at the junction. Indeed, in the case of the 2CC[ɑ], ψ_{10} is very strongly excited ($F_{00,10} > 1.5$) while the other HOM are not excited, and in the case of the 2CE[ɑ], the other HOM are excited and $F_{00,10}$ is close to 0.5 (see Fig. 2.13).



(a) Bandwidths of transfer function H (b) Bandwidths of input impedance Z

Figure 4.2: Bandwidths (at -3 dB) of the formants as a function of the corresponding frequencies of two approximations of the vocal tract shape for the vowel [ɑ] constituted of two cylindrical tubes with either concentric (2CC[ɑ]) or eccentric junctions (2CE[ɑ]).

In the case of the 2CE[ɑ], there are two anti-resonances (at 7 kHz and 7.49 kHz) and four formants (at 6.62 kHz, 7.63 kHz, 8.34 kHz and 9.85 kHz) which are not present in the case of the 2CC[ɑ]. Since, except for the formant at 6.6 kHz, they are above the cut-on frequency of ψ_{01}, they can be understood as the consequence of the excitation and the propagation of this HOM. The fact that there is an additional formant below the cut-on frequency of ψ_{01}, but close to this frequency could mean that the influence of this HOM begins slightly under its cut-on frequency. Indeed, when one gets closer to the cut-on frequency of an HOM, its exponential damping becomes less important and a significant amplitude can be expected at the exit of the geometry.

When it was possible, The -3 dB bandwidths of the formants of H and Z have been extracted. They are presented in Fig. 4.2. Up to the 5 kHz they are almost exactly similar for both H and Z. There are more differences above this frequency, but the same tendency is observed.

The bandwidths of the four first formants are almost similar for both the 2CC[ɑ] and the 2CE[ɑ], and there are more differences above 5 kHz. This can be understood as the consequence of the excitation of ψ_{01} in the case of the 2CE[ɑ].

Except for the fourth formant, the bandwidths of the 2CC[ɑ] increase with the frequency. It is also the case for the formants of the 2CE[ɑ] up to 7 kHz, but there is a sudden decrease above this frequency and it increases again.

This can be explained by the differences in radiation efficiency of ψ_{00} and ψ_{01}. Since the only loss mechanism implemented for these simulations is the radiation outside the geometries, it is hypothesised that the variations of bandwidth can be related to the radiation efficiency, which is quantified by the real part of the radiation impedance. A great value indicates that the losses by radiation are important. In the case of the two tube geometries, $\Re(Z_{00,00})$ is superior to $\Re(Z_{01,01})$ as it can be seen in Fig. 2.6a (the frequency range 0 kHz to 10 kHz corresponds to the range $kR = 0$ to $kR = 2.5$). In the case of the 2CC[ɑ] the sound is radiated by ψ_{00} only and since $\Re(Z_{00,00})$ increases with the frequency, the radiation losses increases

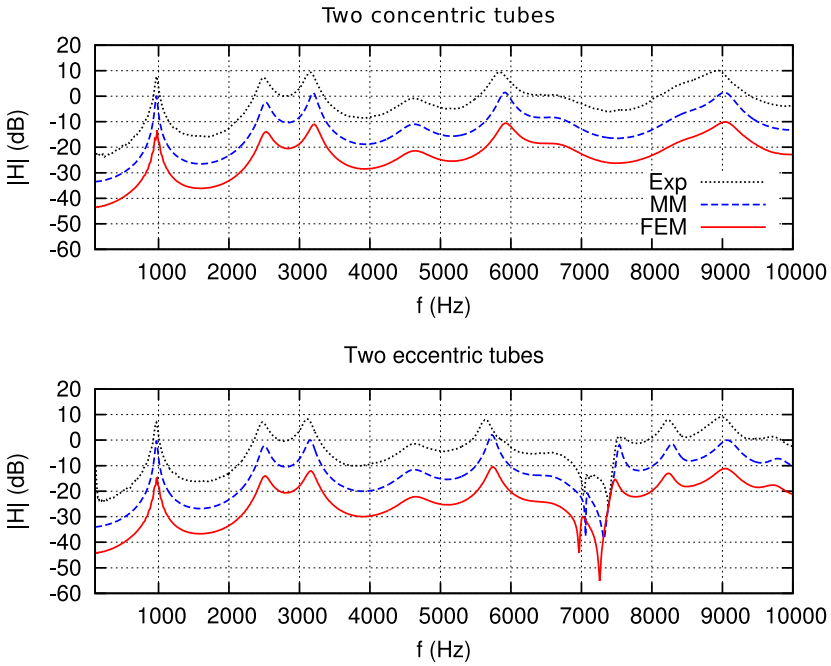


Figure 4.3: Examples of pressure-pressure transfer functions between two points inside two vowel [a] vocal tract replicas made of two cylindrical tubes with (top) concentric (2CC[a]) and (bottom) eccentric junctions (2CE[a]). The Exp and FEM data has respectively been increased and decreased by 10 dB and -10 dB with respect to the MM for visualization purposes. (Exp: experimental data, MM: multimodal method, FEM: finite element method).

which induces wider bandwidths of the formants. Up to the cut-on frequency of ψ_{01} the same reasoning holds for the 2CE[a]. Above this frequency the sound is also radiated by ψ_{01} which has a smaller radiation efficiency. As a consequence, the bandwidths of the formants are reduced and increase again because $\Re(Z_{01,01})$ also increases.

In summary: Below the cut-on frequency of ψ_{01} the acoustical characteristics of 2CC[a] and 2CE[a] are very close. On the contrary, above this frequency significant differences are observed: additional resonances and anti-resonances are observed when ψ_{01} can propagate. The bandwidth of the resonances is also reduced when ψ_{01} is involved in the radiation of the sound, because this HOM has a lower radiation efficiency than ψ_{00}.

4.1.3 Pressure-pressure transfer functions

In order to validate both the MM and FEM simulations, PPTF have been compared. Fig. 4.3 presents examples of PPTF measured and simulated between two points on the propagation axis x_3 (see Fig. 2.1 for x_3 definition). One point is located inside the replica at 30 mm from the source and the other one is outside at 10 mm from the open end. The resonance

Peak frequencies (Hz)										
		plane mode						higher order modes		
		P1	P2	P3	P4	P5	P6	P7	P8	P9
concentric	Exp	970	2490	3160	4620	5840	8930	-	-	-
	MM	970	2520	3190	4630	5910	9030	-	-	-
	FEM	980	2530	3210	4640	5920	9040	-	-	-
eccentric	Exp	970	2480	3110	4630	5640	8990	7530	8230	9720
	MM	970	2500	3150	4630	5730	9070	7540	8280	9790
	FEM	980	2510	3160	4650	5740	9040	7480	8240	9720
Peak bandwidths (Hz)										
concentric	Exp	40	157	159	542	228	422	-	-	-
	MM	31	142	114	482	177	304	-	-	-
	FEM	34	173	153	547	230	386	-	-	-
eccentric	Exp	45	149	156	*	159	316	*	195	*
	MM	32	132	122	505	117	290	65	169	*
	FEM	37	166	160	612	158	353	107	211	*

Table 4.2: Peak frequencies and -3 dB bandwidths of the pressure-pressure transfer functions (PPTF) presented in Fig. 4.3 obtained for two vowel [ɑ] vocal tract replicas made of two cylindrical tubes with concentric (2CC[ɑ]) and eccentric junctions (2CE[ɑ]). (Exp: experimental data, MM: multimodal method, FEM: finite element method *: not measurable).

frequencies and their -3 dB bandwidths are listed in Tab. 4.2.

As expected, above the cut-on frequency of ψ_{01} (6897 kHz at 26.5°C), peaks and dips which are absent in the concentric case, appear for the eccentric one. In experiments, these dips are located at 7030 Hz and 7380 Hz while additional peaks are found at 7530 Hz, 8230 Hz and 9720 Hz (see Tab. 4.2). The peak frequencies of the 2CE[ɑ] slightly differ from the 2CC[ɑ] (up to 3.4 % lower for P5 of the experimental data). Their bandwidths are smaller, up to 30.2 % smaller for P5 of the experimental data. As for the formants of H and Z, the peak above the cut-on frequency of ψ_{01} has a higher frequency in the case of the 2CE[ɑ], and its bandwidth is smaller. The MM and FEM simulations follow the same trend.

Some small differences can also be observed below 5 kHz. As for the formants of H and Z, the peaks mainly occur at a lower frequency for the 2CE[ɑ] compared to the 2CC[ɑ] (up to 1.6 % lower for P3 of the experimental data), except for P1 and P4 which have similar values. This shift down in the peak frequencies could be explained by the fact that the replica behaves as if it is extended by the eccentricity of the junction. Indeed the midline is longer for the eccentric case.

Both MM and FEM peak frequencies are close to the experiments, within less than 2 %. In general, the bandwidths of the simulations are smaller than those obtained experimentally. This can be understood as an effect of visco-thermal losses. Indeed, accounting for visco-thermal losses would reduce the resonance frequencies and increase their bandwidths ([33]),

and the effect is expected to be similar on the peaks of the PPTF. The MM neglects all visco-thermal losses while FEM account partially for them (see Section 3.4) and are thus closer to the experimental values.

In summary: A good agreement between the experimental data and the MM and FEM simulations has been found. Observations similar to the one performed on H and Z have been performed on the PPTF: the same shift down of the peak frequencies at low frequency has been observed and above the cut-on frequency of ψ_{01}, a HOM, higher peak frequencies with lower bandwidths have been observed.

4.1.4 Pressure field maps

In order to investigate the effect of HOM on the internal and the external acoustic field, and to validate both MM and FEM simulation methods, pressure field maps have been measured and simulated.

4.1.4.1 Longitudinal plane pressure field measurement

In Fig. 4.4 is shown as an example the pressure maps for the concentric (top) and eccentric configuration (bottom) when the 7th formant is excited, for both experiments and simulations. Some cuts of these pressure maps are also shown in Fig. 4.5 to compare more easily the measurements and the results obtained with each method.

In agreement with the results observed on the transfer functions, no propagation of the HOM can be appreciated on the pressure field in the concentric case (see top of Fig. 4.4). However, the experimental data exhibit some non planar waveforms (see Fig. 4.4a), in contrast to simulations (see Fig. 4.4b and Fig. 4.4c). Although the greatest care was taken when building the replica, it appears that it is not perfectly symmetrical. As a consequence the first HOM, ψ_{01}, is slightly excited and propagates. The influence of degree of eccentricity of a junction is further investigated in section 5.2.1. However, the longitudinal pressure profile along x_3 is located on the nodal line of ψ_{01} and allows one to observe the effect of the plane mode alone. The measured profile appears to be similar to the one obtained in simulations (see Fig. 4.5a). Secondly, the transverse profile along x_2 presented in Fig. 4.5b shows that this effect has a limited impact on the radiated pressure. On the other hand, the iso-amplitude lines are almost circular outside of the replica (see top of Fig. 4.4 and Fig. 4.5b), showing that the acoustic pressure is radiated in almost spherical waves.

In what concerns the 2CE[a], the first HOM ψ_{01} effects can be observed in both experiments and simulations (see bottom of Fig. 4.4 and Fig. 4.5). While the plane mode is still predominant in the narrowest tube, ψ_{01} becomes predominant in the widest cross-section of the replica. The influence of this mode on the radiated field can be observed as well. The radiated sound pattern is no longer spherical (see bottom of Fig. 4.4 and Fig. 4.5d). This is

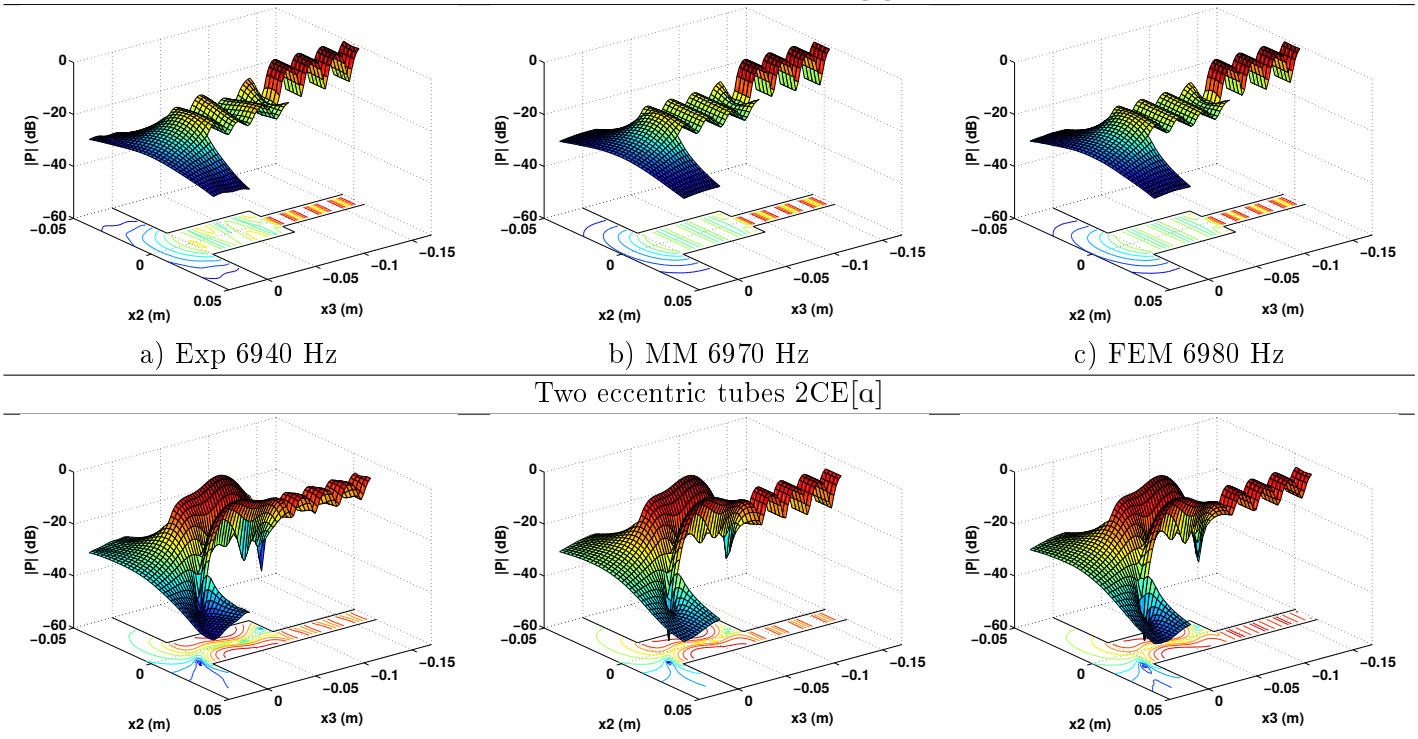
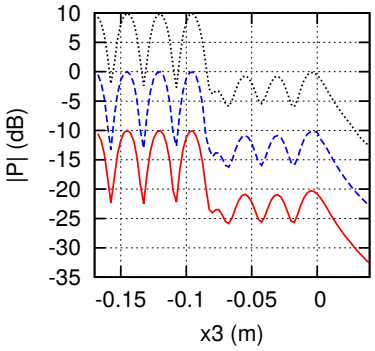
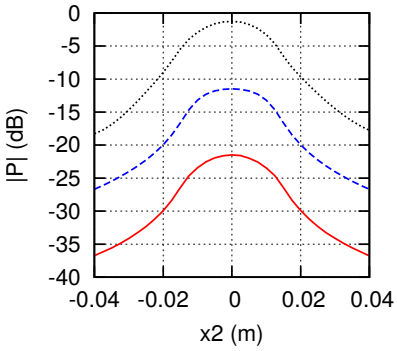


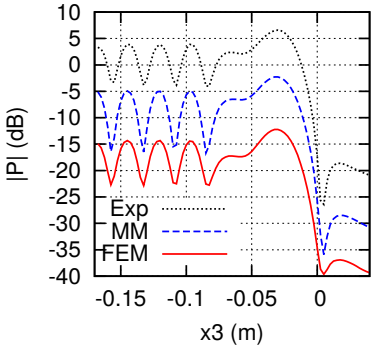
Figure 4.4: Examples of pressure amplitude maps inside and outside of two vowel [a] vocal tract replicas made of two cylindrical tubes with a concentric (2CC[a] top) and an eccentric (2CE[a] bottom) junction for high frequencies (Exp: experimental data, MM: multimodal method, FEM: finite element method, here the origin of the coordinate system is located in the center of the exit plane).



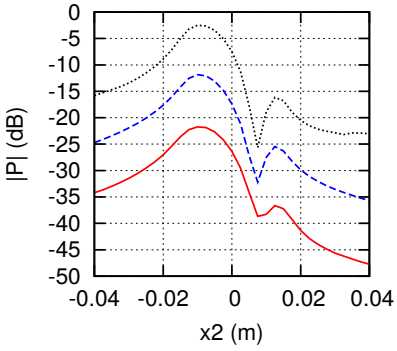
a) x_3 cut two concentric tubes (2CC[a])



b) x_2 cut two concentric tubes (2CC[a])



c) x_3 cut two eccentric tubes (2CE[a])



d) x_2 cut two eccentric tubes (2CE[a])

Figure 4.5: Cuts of the pressure maps in Fig. 4.4 along the central axis x_3 (left column) and a perpendicular x_2 axis located just in front of the open end (right column) measured and simulated on two vowel [a] vocal tract replicas made of two cylindrical tubes with either concentric or eccentric junctions (2CC[a] and 2CE[a]). A 10 dB shift has been applied between the maxima of the three curves for visualization purposes. (Exp: experimental data, MM: multimodal method, FEM: finite element method).

investigated in further detail in section 5.1.

Finally, the different curves shown in Fig. 4.5 have been quantitatively compared. The mean difference between the experiment and the simulations is less than 1 dB for the concentric case and less than 2 dB for the eccentric case.

In summary: As for the PPTF, a good agreement has been found between the experimental data and the MM and FEM simulations. The observation of the pressure field showed that, as expected, the plane wave assumption is not valid above the cut-on frequency of ψ_{01} when ψ_{01} is excited, and that the radiation of the sound is affected by the propagation of ψ_{01}. On the other hand it has been observed that the excitation of ψ_{01} is very sensitive to small asymmetries.

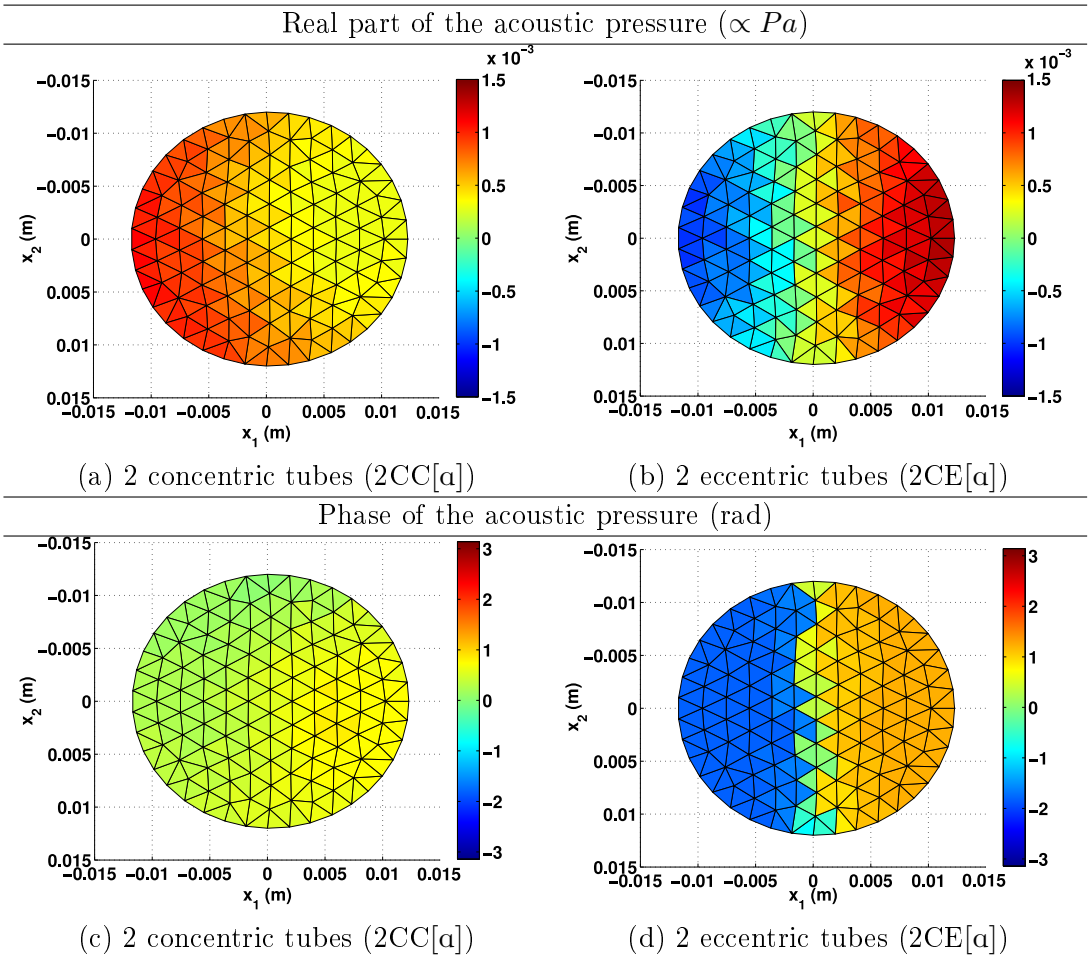
4.1.4.2 Transverse plane (x_1,x_2) pressure field measurement

In the purpose of identifying the propagation modes involved and to quantify their contribution, pressure field measurements have been performed in the plane (x_1,x_2) perpendicular to the propagation axis x_3. In order to have a distribution of the measurement points as homogeneous as possible, it has been chosen to place them on the nodes of a triangular mesh instead of a square one. Indeed, the edges of a circular shape are more evenly covered using this type of distribution. The real part and the phase of the measured pressure field are presented in Fig. 4.6. The real part of acoustic pressure corresponds to the magnitude of the pressure at the time instant corresponding to $\phi_m - \phi_s + \omega t = n\pi$, n being a relative integer, which yields $t = (n\pi - \phi_m + \phi_s)/\omega$. The magnitude of the real part is thus not the amplitude, and can be smaller. However, this quantity is interesting to study in this context because, unlike the amplitude, it allows one to see the sign change of the acoustic pressure.

It has been chosen to investigate the pressure field at a frequency above the cut-on frequency of ψ_{01} at 7.4 kHz inside the 2CC[a] at 1 cm after the junction inside the larger tube. Indeed, close to the junction the amplitude of the evanescent modes is non-negligible and investigating their amplitude can help to confirm the hypothesis formulated earlier in section 4.1.2. A similar measurement has been performed in the 2CE[a] at the frequency corresponding to the first anti-resonance, 7.06 kHz. Indeed, at this frequency the amplitude of ψ_{01} is expected to be particularly high.

In the case of the 2CC[a], one can see in Figs. 4.6a and 4.6c that the variations of the real part and the phase of the acoustic pressure are small. On the other hand, the real part and the phase have the same sign everywhere on the surface. Thus, as expected, one can see that even above the cut-on frequency of the first HOM, ψ_{01}, the plane mode ψ_{00} is predominant.

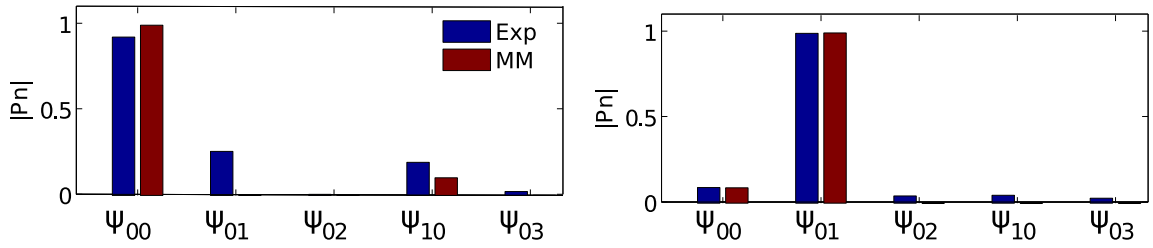
However, the amplitude of the variations of the real part and the phase over the surface is not negligible: $0.0008 \propto Pa$ and 0.7 rad. One can also see that the pressure field is divided into two regions: the left hand side has a higher real part and a lower phase than the right hand side. This can be explained by the fact that the HOM amplitude is small, but not zero.



(a) 2 concentric tubes (2CC[ɑ]) (b) 2 eccentric tubes (2CE[ɑ])

(c) 2 concentric tubes (2CC[ɑ]) (d) 2 eccentric tubes (2CE[ɑ])

Figure 4.6: Real part and phase of the acoustic pressure measured in the plane (x_1,x_2) perpendicularly to the propagation axis x_3, 4.5 cm after the junction inside larger part of the two tube replicas of the vowel [ɑ] having either a concentric (2CC[ɑ]) or an eccentric (2CE[ɑ]) junction. The real part is proportional to Pa to the sensitivity of the microphone and the phase is computed as the difference between the phase of the microphone and the phase of the sound source input signal.



(a) 2 concentric tubes (2CC[a]) (b) 2 eccentric tubes (2CE[a])

Figure 4.7: Modale amplitudes $|P_n|$ of the pressure measured and simulated on a plane perpendicular to the propagation axis of two tube approximations of the vowel [a] with either concentric (2CC[a]) or eccentric (2CE[a]) junctions.

It is hypothesised that this comes from the fact that the replica is not perfectly axisymmetric and that ψ_{01} is slightly excited.

Unlike the 2CC[a], the acoustic pressure has consequent variations over the surface in the case of the 2CE[a]. The pressure field is clearly divided into two regions with opposite sign separated by a line which coincides with the axis x_2. The amplitude of the variation of the real part and the phase is larger than for the 2CC[a]: 0.0023 $\propto Pa$ and 3.1 rad. Moreover, the phase difference between the two regions is almost π, so one can consider that these two regions have opposite phase. The mode-shape of ψ_{01} (see Fig. 2.2) can be clearly recognized. Thus, one can conclude that, as expected, above its cut-on frequency, ψ_{01} can be predominant in the 2CE[a] geometry.

However, the averaged values of the real part and the phase are not zero: 0.00022 $\propto Pa$ and -0.25 rad. On the other hand the real part values are higher on the right hand side. This can be understood as the effect of the plane mode, which is still present though its amplitude is small.

In order to investigate more precisely and more quantitatively the contributions of the different propagation modes to the acoustic field, the measured pressure fields have been projected on the basis of the propagation modes corresponding to a circular shape (see section 2.1.1.1) using the method detailed in section 3.3.7. This allowed one to compute their amplitudes P_n which are presented in Fig. 4.7. These amplitudes have been compared with the one theoretically computed with the MM. In order to make this comparison possible, they have been normalized by the sum of the square of the amplitudes P_n^2 of each mode, $\boldsymbol{P}^\dagger\boldsymbol{P}$. Thus, they are necessarily smaller than one.

As expected from the theory and the observation of the Figs. 4.6a and 4.6c, one can see on Fig. 4.7a that in the case of the 2CC[a] the plane mode ψ_{00} has the highest amplitude. The assumption that ψ_{01} is slightly excited in the experiment is also confirmed: unlike the MM, the experiment shows a non zero amplitude for ψ_{01}. On the other hand, the amplitude of ψ_{00} is smaller for the experiment, which is consistent with the fact that a part of the energy is carried by ψ_{01} in this case. It is hypothesised that this difference comes from the fact that, even though it was carefully built, the replica is not perfectly axisymmetric. One can also

notice that the first axisymmetric HOM ψ_{10} has a non zero amplitude for both the experiment and the MM. Since its cut-on frequency (14354 Hz at 26.5°C) is higher than the frequency of these experiments and simulations, one can conclude that it is an evanescent mode generated by the discontinuity corresponding to the junction between the two tubes. This observation is consistent with the hypothesis formulated in section 4.1.2 from the projection matrix F that ψ_{10} is more strongly excited in the 2CC[a]. The fact that its amplitude is almost two times higher in the experiment, could mean that its exponential decay is overestimated in the MM, but maybe the small imperfections could also play a role in this difference. The agreement between MM and experiment is not so good: the relative difference $\left(\frac{|P_{MM}-P_{Exp}|}{P_{Exp}}\right)$ is of 7.5% for ψ_{00} and 47% for ψ_{10}.

The assumptions formulated for the pressure field observed in the 2CE[a] are also confirmed. Indeed, one can see in Fig. 4.7b that the mode ψ_{01} has the highest amplitude and predominates. On the other hand, the plane mode amplitude is non-zero. The amplitude obtained for ψ_{00} and ψ_{01} with the MM simulation and the experiment is closer than for the 2CC[a]: the relative difference is of 1.9% for ψ_{00} and 0.26% for ψ_{01}. This better agreement can be explained by the fact that there is no ambiguity concerning the axisymmetry and the excitation of ψ_{00} and ψ_{01}. The axisymmetric case would thus be a theoretical case difficult to reproduce accurately experimentally. The amplitude of the HOM ψ_{02}, ψ_{10} and ψ_{03} is small but non zero in the case of the experiment, as for the 2CC[a], the origin of this small discrepancy can be attributed to inaccuracies of the replica.

In summary: The measurements and the simulations in the plane (x_1,x_2) perpendicular to the propagation axis x_3 allowed one to confirm more directly the hypothesis formulated concerning the excitation of ψ_{01} and of the evanescent HOM. The modal projection have proved to be an efficient tool for estimating the modal amplitude P_n of the propagation modes. With the exception of the discrepancies generated by the small imperfections of the 2CC[a], a good agreement has been found between the theoretical modal amplitudes and the one provided by the MM.

Conclusions concerning the two tube geometries: The two tubes replicas point out the evidence for HOM in a simplified two tubes vocal geometries. The importance of the connection of the junctions appears to be critical. When present, these HOM strongly affect both the internal and the radiated sound field. In the next section is evaluated if these findings still hold for more complex vocal tract approximations.

4.2 44 tubes approximations

In order to study the effect of HOM on more realistic geometries, 44 tubes approximations of the vocal tract geometry for vowel [a] have been built from the area function provided by Story [23]. These approximations correspond to the state of the art of the geometrical vocal

tract approximation used in speech studies. Their description with an area function gives a lot of freedom for the precise realisation of the geometries. Indeed, the cross-sectional shape and the eccentricity of the junction can be varied and compared while the area function used to design it remains similar. This allows one to test a variety of parameters and to evaluate their influence on the vocal tract transfer function and the directivity of the radiated sound.

In order to perform measurements on these geometries, the three vowel [a] vocal tract replicas presented in Figs. 3.1d, 3.1e and 3.1f of Section 3.1 have been used. In this case the measurements have been limited to the frequency range (2kHz - 10kHz), of more interest for the purposes of this work. Only one sound source was required for this frequency range, and the measurement procedure was thus more simple.

4.2.1 Propagation modes inside the geometries

Though more complex, the 44 tube geometries have the same kind of shape as the two tubes ones: a narrow part (from x3 = 0 mm to x3 = 80 mm, see Fig. 3.2) is followed by a wider one (from x3 = 80 mm to x3 = 167 mm). Thus, for the same reason as for the two tube geometries, only the plane mode ψ_{00} is expected to propagate inside the 44CC[a], and only ψ_{00} and ψ_{01} are expected to propagate inside the 44CE[a].

In the case of the 44EE[a], because of the elliptical cross-sectional shape, the propagation modes involved have different mode-shapes (see Fig. 2.4). However, the area function remains similar, and a narrow part in which only plane waves can propagate is connected to a wider part in which HOM can propagate. Thus, as for the 44CE[a], the HOM are excited at the entrance of this wider part by the coupling with the plane mode ψ_{e01}. Because their contributions compensate each other in Eq. (2.36), all the HOM featuring a vertical nodal line on the vertical symmetry axis x_2 are not expected to propagate. So the only HOM expected to propagate between 0 kHz and 10 kHz are ψ_{e21} and ψ_{o11}. However, ψ_{o11} is expected to have a very limited influence on this frequency range since its cut-on frequency, 9860 Hz at 22.65°C in the widest part, is very close to 10 kHz. Unlike the eccentric configurations with circular cross-sections (2CE[a] and 44CE[a]), the eccentricity does not play an important role in the excitation of ψ_{e21}. Indeed, this HOM is excited because a larger area in the center is covered by the intersection with a smaller cross-section than the areas with opposite phase on the edges. As a consequence, this HOM can be excited with an axisymmetric configuration.

4.2.2 Simulated velocity-pressure transfer functions and input impedance

As for the two tube geometries, the input impedance Z and the transfer function H have been computed with the MM. The correspondence between the peaks of Z and H, with a difference smaller than 0.7%, is better than for the two tube geometries. It is interesting to notice that, unlike for the 2CE[a], the anti-resonances do not necessarily correspond to a dip of input impedance.

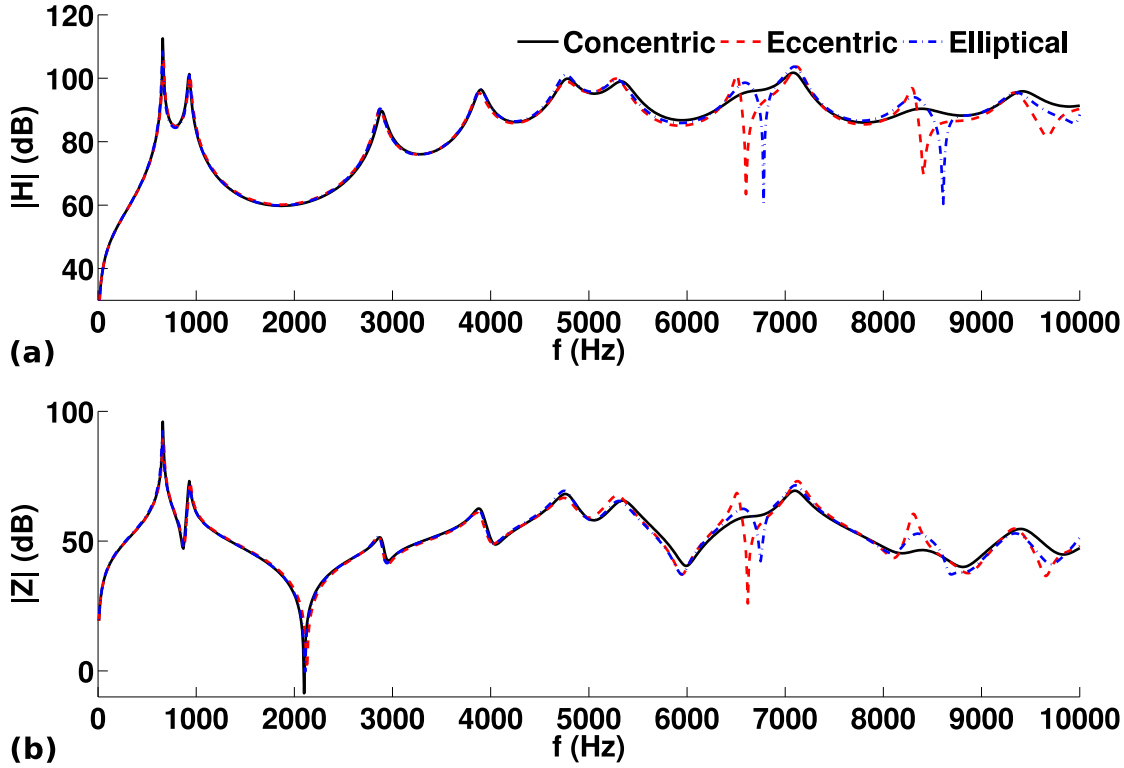


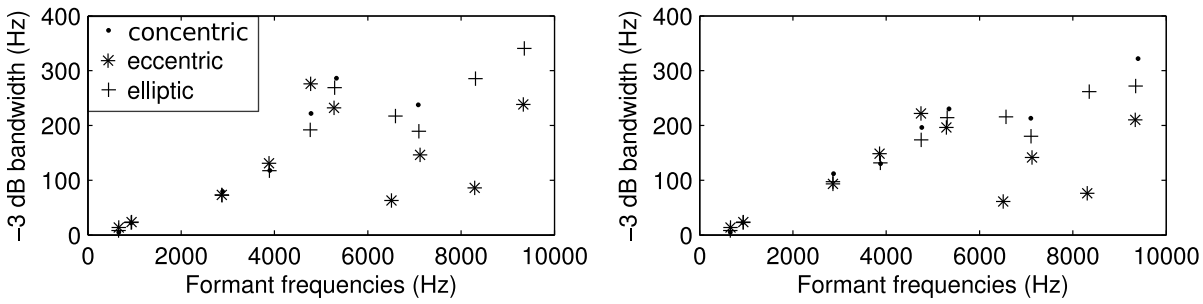
Figure 4.8: Amplitude of the simulated (with MM) input impedance $|Z|$ and of the transfer function $|H|$ between the input volume velocity and the acoustic pressure radiated at 60 cm in front of the exit of three vocal tract approximations for the vowel [ɑ] based on ([23]) with concentric junction and circular cross-section, eccentric junctions and circular cross-sections and eccentric junctions in the plane (x_2,x_3) and elliptical junctions.

The frequencies of the two first formants are 660 Hz and 930 Hz, which is similar to the values given in [23]. As for the two tube geometries, these values do not depend on the junction type. If there is a difference of only 10 Hz with the frequency of the first formant of the two tube geometries (670 Hz), the difference with the second formant is much larger (1200 Hz for the 2CC[a] and the 2CE[a]). This can be explained by the fact that the two tube shape has been designed so that the frequencies of the two first formants corresponds to the averaged values observed for the vowel [a] [1], whereas the area function of the 44 tube geometries has been determined from MRI measurements [23]. In this case, the perturbations related to the measurement process (supine position in a noisy environment, see [23] for a detailed discussion) can induce formant frequencies different from the averaged value, but which still corresponds to a vowel [a].

The frequencies of the formants of 44CC[a] are globally higher than the one of 44CE[a] up to 1% higher for the sixth formant. However, the first, second and seventh formant are lower, down to 0.9% for the second. The same tendency is observed between the 44CC[a] and the 44EE[a]: with the exception of the second and the seventh formants, their frequency is higher for the 44CC[a] (up to 0.9% for the eighth). So the reduction of the frequency of the formants under the cut-on frequency of the first propagating HOM (5885 Hz for ψ_{01} for the 44CE[a] at 23.2°C and 6084 Hz for ψ_{e21} for the 44EE[a] at 22.65°C) when eccentric junctions are introduced is not as clearly observed for the 44 tube geometries as for two tube ones. This can be understood as the consequence of the correction of the length of the midline (see section 3.1) which has been successful in limiting the variations of the formants.

Up to 6 kHz the input impedance curves of the three configurations are almost exactly similar, and differences are observed above this frequency. However, unlike the two tube geometries, the differences are limited to three frequency intervals: 6.2 kHz to 7 kHz, 7.9 kHz to 8.6 kHz and 9.3 kHz to 9.8 kHz. In addition, no significant change of the global amplitude of H is observed. This could be interpreted as the fact that the variations of cross-sectional area are more progressive in this case. As for the two tube geometries, additional formants and anti-resonances are observed when eccentric junctions are used and when the cross-section shape is changed. An additional formant (at 6.51 kHz) and three anti-resonances (at 6.6 kHz, 8.41 kHz and 9.67 kHz) can be seen in the transfer function of the 44CE[a]. In the case of the 44EE[a], there is also an additional formant (at 6.59 kHz) and three anti-resonances (at 6.78 kHz, 8.61 kHz and 9.95 kHz).

The frequencies of the formants of the 44CE[a] and the 44EE[a] are quite close: the maximal difference is of 1.3% for the seventh formant which is higher in the case of the 44EE[a]. There are more differences in the frequency of the anti-resonances, up to 2.9% higher for the second anti-resonance of the 44EE[a]. This can be explained by the fact that the formants which are in common with the 44CC[a] are mainly due to the propagation of the plane mode and that the anti-resonances and the additional formant are due to the propagation of HOM which are different in the case of the 44CE[a] and of the 44EE[a]. This hypothesis is reinforced by the fact that the difference between the frequency of the formants of the 44CE[a] and the 44EE[a] is maximal for the additional formant. The fact that the additional formant and the anti-resonances are higher for the 44EE[a] can be related to the fact that the HOM ψ_{e21}



(a) Bandwidths of transfer function H (b) Bandwidths of input impedance Z

Figure 4.9: Bandwidths as a function of the corresponding formants frequencies of three approximations of the vocal tract shape for the vowel [ɑ] constituted of 44 tubes with circular cross-section and concentric junctions (44CC[ɑ]), circular cross-section and eccentric junctions (44CE[ɑ]) and elliptical cross-sections and eccentric junctions in the plane (x_2,x_3) (2ET[A]).

involved in this geometry has a cut-on frequency (6084 Hz) higher than the one of ψ_{01} (5885 Hz) which is involved in the 44CE[ɑ].

The -3 dB bandwidths of the formants of H and Z have been extracted and are presented in Fig. 4.9 as a function of the formant frequency. Since the only loss mechanism implemented for these simulations is the radiation outside the geometries, it is hypothesised that the variations of bandwidth can be related to the radiation efficiency which is quantified by the real part of the radiation impedance. The same evolution is observed for H and Z. For all the geometries it increases with the frequency up to 6 kHz, above which it varies and more differences can be seen between the geometries. This global increase can be explained by the fact that the losses by radiation increases with the frequency. Indeed, the real part of the radiation impedance of the plane mode ψ_{00} increases in the frequency range 0 kHz to 10 kHz.

The bandwidths of the 44CE[ɑ] are globally wider than the one of the 44EE[ɑ] between 0 kHz and 5 kHz, up to 84 Hz wider for the fifth formant. The opposite is observed above this frequency: the bandwidths of the 44CE[ɑ] are smaller than the ones of the 44EE[ɑ], down to 200 Hz for the ninth formant. This can be explained by the fact that up to the fifth formant ψ_{00} is predominant. The real part of its radiation impedance is smaller with an elliptical cross-section than a circular one (considering the same cross-sectional area). Thus, the radiation losses are less important with elliptical cross-section and the bandwidth of the formants tends to be smaller in this frequency range. Above 5 kHz the influence of the HOM is more important. In the case of the 44CE[ɑ], for all the junctions, the coupling term between ψ_{00} and ψ_{01} is higher than the coupling term between ψ_{e01} and ψ_{e21} in the case of the 44EE[ɑ]. Indeed, as an example, $F_{00,01} = 0.52$ at the junction between the section 22 and 23, which corresponds to the beginning of the wide part, and $F_{e01,e21} = 0.4$. The mode ψ_{e21} is thus less strongly excited and its contribution to the radiation is less important than ψ_{01} in the case of the 44CE[ɑ]. Given that the real part of the radiation impedance of ψ_{01} and ψ_{e21} are smaller than the one of ψ_{00} and ψ_{e01}, the radiation efficiency of the 44CE[ɑ] is reduced compared to the 44EE[ɑ] and its bandwidths are smaller.

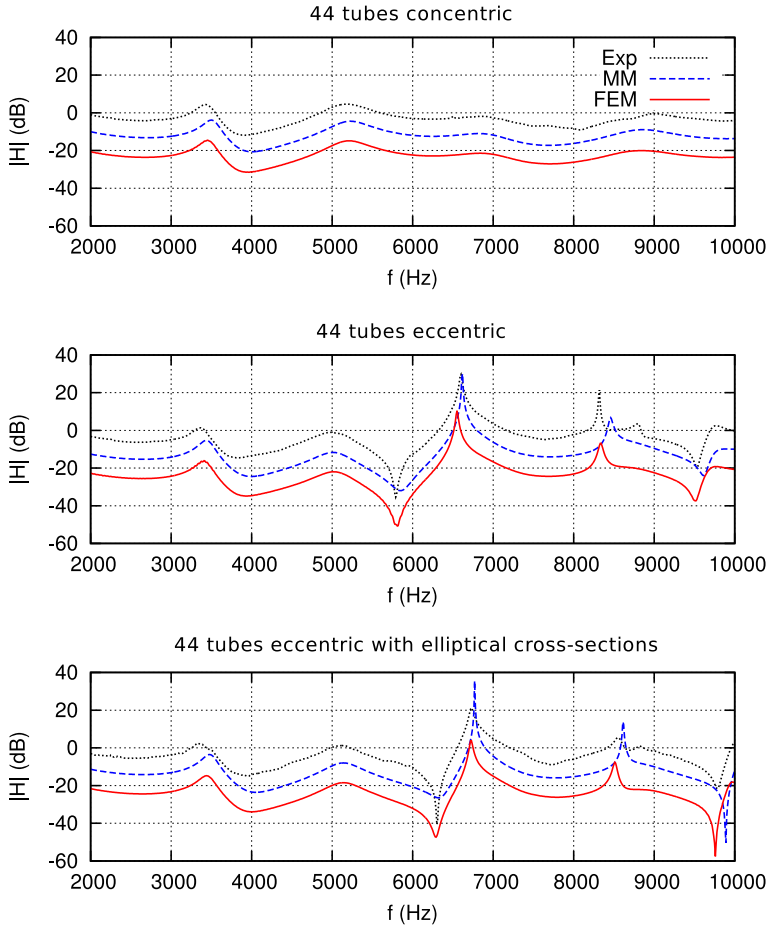


Figure 4.10: Examples of pressure-pressure transfer functions (PPTF) between two points inside three simplified vowel [ɑ] vocal tract replicas with (top) concentric junctions and circular cross-sections (44CC[ɑ]), (middle) eccentric junctions and circular cross-sections (44CE[ɑ]) and (bottom) eccentric junctions and elliptical cross-sections (44EE[ɑ]). (Exp: experimental data, MM: multimodal method, FEM: finite element method).

4.2.3 Pressure-pressure transfer functions

In order to validate the MM and FEM simulation methods for the 44 tube geometries in the frequency domain, PPTF have been simulated and measured. PPTF between two points inside the replicas located at 80 mm and 130 mm from the source on x_3 are presented in Fig. 4.10. The peaks and dips frequencies and -3 dB bandwidths of the peaks of these transfer functions have been listed in Tab. 4.3.

As expected from Section 4.1, one can observe peaks (P3 and P4) and dips (D1 and D2) for the geometries which feature eccentric junctions and elliptical cross-sections (44CE[ɑ] and 44EE[ɑ]) which are not observed for the concentric configuration (44CC[ɑ]).

In the mid-frequency range (2 kHz - 5.5 kHz) the peaks occur at lower frequencies for the

Peak and dip frequencies (Hz)							
		plane mode		HOM		dips	
		P1	P2	P3	P4	D1	D2
concentric	Exp	3430	5180	-	-	-	-
	MM	3500	5230	-	-	-	-
	FEM	3450	5210	-	-	-	-
eccentric	Exp	3360	4970	6600	8320	5790	9530
	MM	3440	5010	6620	8460	5850	9610
	FEM	3410	5020	650	8340	5810	9510
elliptic	Exp	3340	5120	6730	8550	6300	9790
	MM	3480	5140	6770	8610	6320	9890
	FEM	3440	5140	6730	8510	6290	9790
Peak and dip bandwidths (Hz)							
concentric	Exp	227	552	-	-	-	-
	MM	208	512	-	-	-	-
	FEM	213	507	-	-	-	-
eccentric	Exp	210	478	31	14	30	65
	MM	190	435	14	49	211	102
	FEM	183	446	27	55	72	86
elliptic	Exp	154	433	54	113	15	72
	MM	184	433	4	17	258	6
	FEM	205	454	41	48	64	8

Table 4.3: Peaks and dips frequencies and bandwidths of examples of transfer functions (presented in Fig. 4.10) between two points inside of three vowel [ɑ] vocal tract replicas with concentric junctions and circular cross-sections (44CC[ɑ]), eccentric junctions and circular cross-sections (44CE[ɑ]) and eccentric junctions and elliptical cross-sections (44EE[ɑ]). The Exp and FEM data have, respectively, been increased and decreased by 10 dB and -10 dB with respect to the MM for visualization purposes. (Exp: experimental data, MM: multimodal method, FEM: finite element method).

44CE[a] and the 44EE[a]. Indeed, for the experimental data they are up to 4% lower (for P2) for the 44CE[a] and the 44EE[a]. So, the same tendency as the one observed for the resonances of H and Z is observed here with a larger magnitude (the difference is up to 1% for H and Z). Another difference between the 44CC[a] and the other geometries at the mid-frequencies is the reduction of the peak bandwidths. This effect is stronger than on P2 and P3 of the two tubes replicas. Indeed, for the 44CE[a] the bandwidths of P1 and P2 are 7.3% and 13.5% smaller than for the 44CC[a]. The effect is more important for the 44EE[a]: P1 and P2 are 32.3% and 21.4% smaller than for the 44CC[a]. This tendency is only observed for the third and the sixth formant of H and Z. The same trend is observed on simulations, however, the effect is more pronounced for the 44CE[a] on FEM simulations. Thus, it appears that the effect of the eccentricity of the junctions and the cross-sectional shape on the bandwidths of the peaks is not negligible, even at mid-frequencies.

The PPTF obtained for the 44EE[a] is similar to the one obtained for the 44CE[a] up to the high frequencies where the peak and dip frequencies are higher for the 44EE[a] (see middle and bottom of Fig. 4.10). Indeed, for the 44EE[a] the frequency of P4 and D1 are respectively 230 Hz and 510 Hz higher (or 2.8 % and 8.8 % higher). This can be related to the higher resonances and anti-resonance frequencies observed for the 44EE[a] in H and Z. Similarly, it can be related to the fact that due to the difference of cross-section, different HOM having different cut-on frequencies are involved in the 44CE[a] and the 44EE[a]. As for H and Z, the bandwiths of the 44CE[a] are wider than the ones of the 44EE[a] below the cut-on frequency of the first propagating HOM, up to 56 Hz higher for P1. And, likewise, the opposite is observed above this frequency: the bandwidths of the 44CE[a] are smaller than the ones of the 44EE[a], down to 99 Hz smaller for P4.

Both MM and FEM simulations are in good agreement with the experiments. The maximal difference between peaks and dips frequencies is less than 5 %. The bandwidths of the simulations are smaller than in the experiments. The difference is more pronounced for the MM. As for the two tubes replicas this can be attributed to the visco-thermal losses which are partially taken into account by the FEM (see Section 3.4) but neglected by the MM. On the other hand, on the experimental data of Fig. 4.10 one can notice an additional peak above P4 for the 44CE[a] and the 44EE[a] which does not appear in the simulations. The presence of this peak may be due to experimental issues such as positioning error or small imperfections in the replica.

In summary: Similar differences as the one observed between the 2CC[a] and the 2CE[a] are observed between the 44CC[a] and the 44CE[a]. Likewise, additional peaks and dips are observed in the case of the 44CE[a] and the 44EE[a]. The difference of cross-sectional shape has been observed to influence the frequency and the bandwidth of the peaks and dips of the PPTF. As for the PPTF of the two tube geometries, similar characteristics have been found between the PPTF and H and Z. Thus, higher peak and dip frequencies and bandwidths are observed above the cut-on frequency of the first propagating HOM for the 44EE[a] than for the 44CE[a]. Eventually, a good agreement has been found between the simulations and the experiment.

4.2.4 Pressure field maps

In order observe the effect of HOM on the pressure field and to validate the MM and FEM simulation methods for the 44 tube geometries, pressure maps have been simulated and measured. In Fig. 4.11 examples of pressure maps are presented for the 44CC[a] (top), the 44CE[a] (middle) and the 44EE[a] (bottom) at high frequency values for both experiments and simulations. Longitudinal and transversal cuts of these maps along x_3 and x_2 at 2.5 mm from the open end outside of the replicas are provided in Fig. 4.12.

For the 44CC[a] one can observe the effect of the HOM inside the replica for the experimental data (see Fig. 4.11a) for which significant variations of the amplitude along x_2 can be seen. This is not the case for the simulated data for which the iso-amplitude lines are perpendicular to x_3 inside the replicas (see Fig. 4.11b and Fig. 4.11c). As for the two tube geometries, it is hypothesised that this is due to small imperfections of the replica. The pressure profile along the central axis x_3 provided in Fig. 4.12a allows one to observe the effect of the plane mode alone. It appears to be similar to the simulation profiles. On the other hand, the transverse pressure profile along x_2 presented in Fig. 4.12b shows that the effect on the radiated pressure is limited. As a matter of fact, the iso-amplitude lines of the radiated pressure are almost circular.

As expected, this is not the case for the 44CE[a] and the 44EE[a]. The effect of the HOM inside these replicas can be clearly observed (see middle and bottom of Fig. 4.11). Significant variations of the amplitude along x_2 are observed and there are local maxima and minima of the amplitude inside the geometries. The minima obtained with MM and FEM are more pronounced than those measured experimentally, and the minima of the MM simulations are more important than those of the FEM simulation (see middle and bottom of Fig. 4.11 and Fig. 4.12). This can be related to the observations made on the PPTF: the peaks of the simulations have higher amplitudes and narrower bandwidths. The origin of this difference is probably also the poor representation of visco-thermal losses in the simulations.

The pressure radiated by the 44CE[a] and the 44EE[a] appears as strongly directional though it is less noticeable for the 44EE[a] (see middle and bottom of Fig. 4.11).

In Fig. 4.12 both MM and FEM simulations show pressure profiles which are close to the experimental data. There are however some small differences in the amplitude of the minima and the symmetry of the patterns in the cuts performed along x_2 outside of the vocal tracts (see Fig. 4.12b, Fig. 4.12d and Fig. 4.12f). In Fig. 4.12b and 4.12f the experimental pressure profile is not perfectly symmetrical whereas the simulation pressure profiles are. In Fig. 4.12d the two lobes of the experimental data have almost the same amplitude whereas for both MM and FEM the left lobe has a smaller amplitude than the right lobe. Small differences in the amplitude of the minima can also be observed on the cuts performed along x_3 (see Fig. 4.12a, Fig. 4.12c and Fig. 4.12e). It is assumed that the difference comes from an experimental artifact due either to some imperfections of the replicas or to reflections on the table which has been insufficiently damped by the acoustic foam.

To compare more directly the cuts of Fig. 4.12 the same procedure as for the two tubes

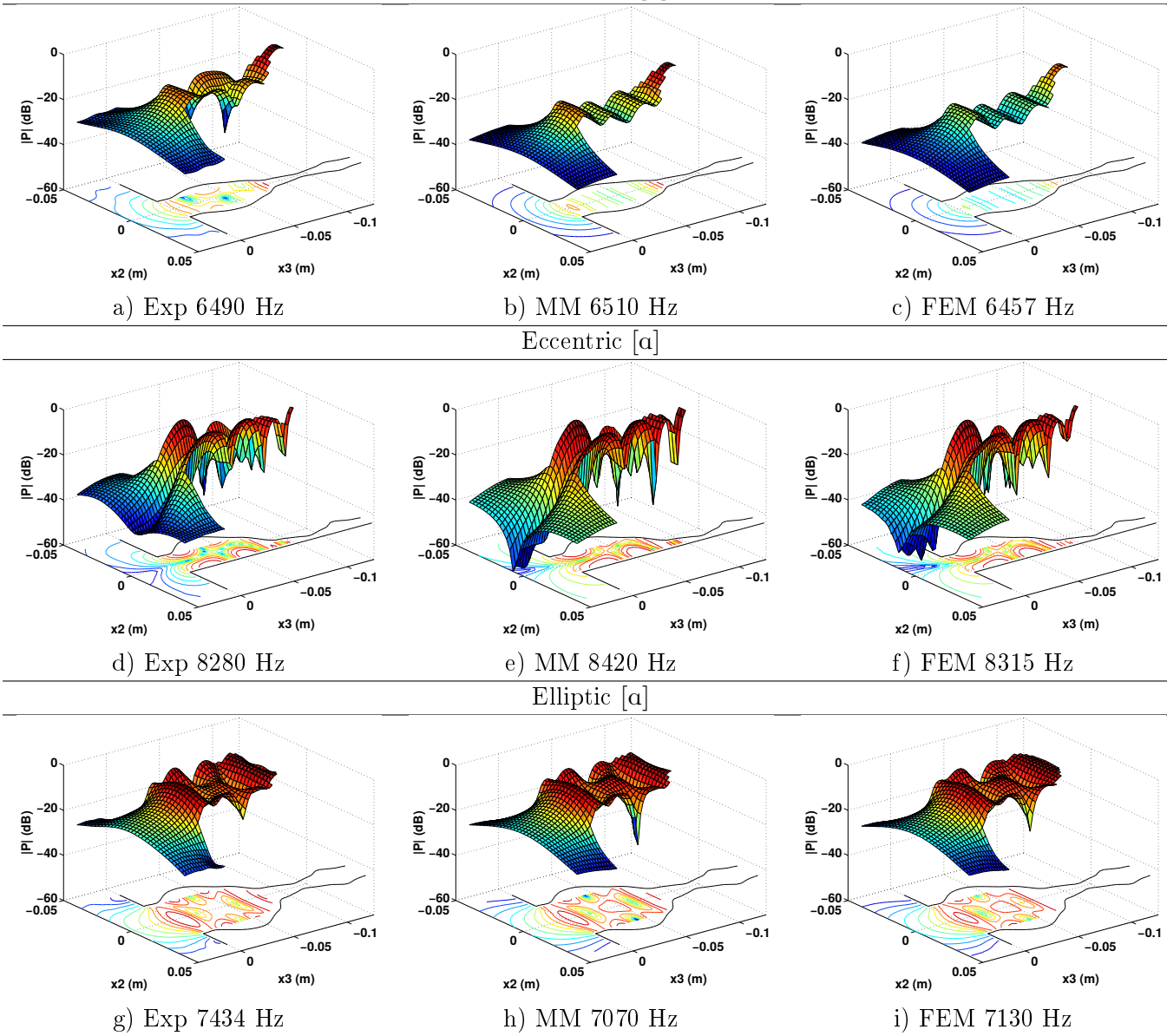
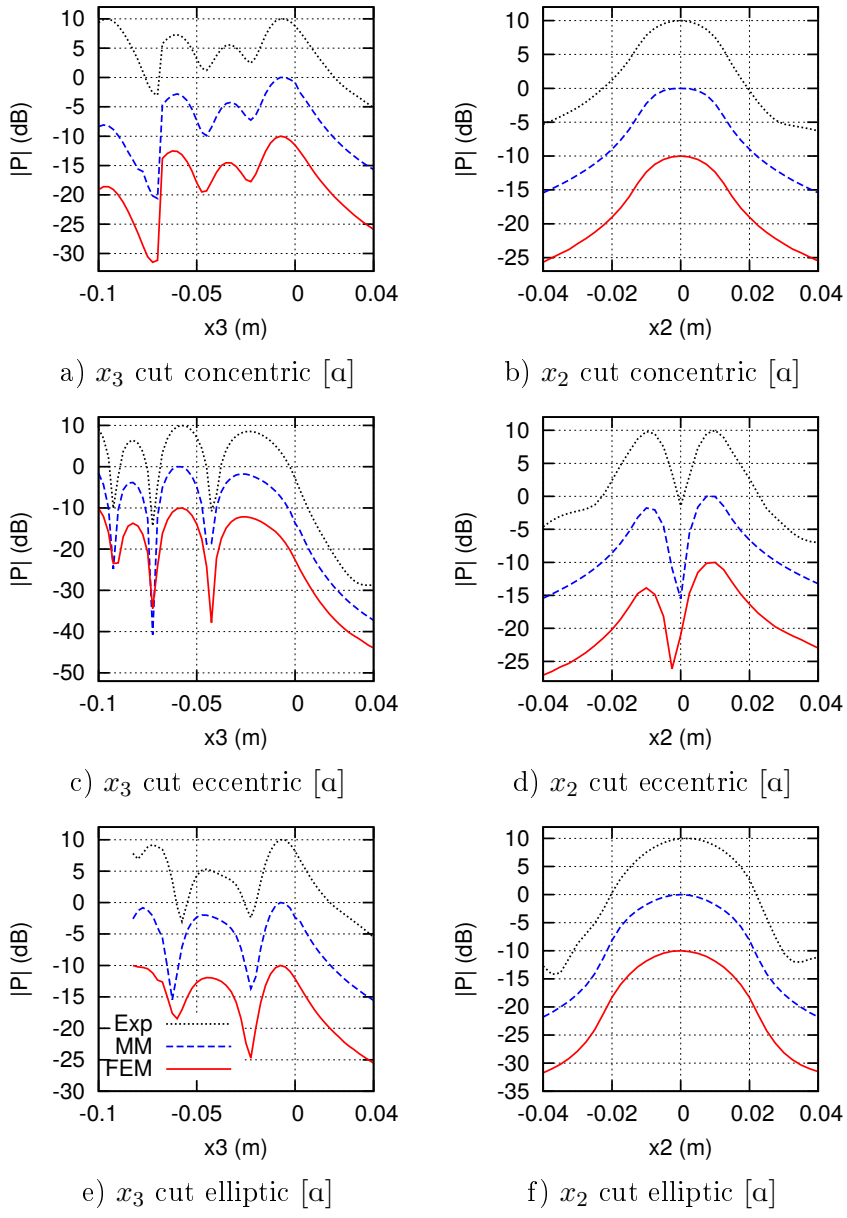


Figure 4.11: Examples of pressure amplitude maps for high frequencies inside and outside of three simplified vowel [a] vocal tract replicas with (top) concentric junctions and circular cross-sections, (middle) eccentric junctions and circular cross-sections and (bottom) eccentric junctions and elliptical cross-sections. (Exp: experimental data, MM: multimodal method, FEM: finite element method, here the origin of the coordinate system is located in the center of the exit plane).



a) x_3 cut concentric [a] b) x_2 cut concentric [a]

c) x_3 cut eccentric [a] d) x_2 cut eccentric [a]

e) x_3 cut elliptic [a] f) x_2 cut elliptic [a]

Figure 4.12: Cuts of the pressure maps in Fig. 4.11 along the propagation axis x_3 (left column) and the axis x_2 just in front of the open end (right column) measured and simulated on three simplified vowel [a] vocal tract replicas with (top) concentric junctions and circular cross-sections, (middle) eccentric junctions and circular cross-sections and (bottom) eccentric junctions and elliptical cross-sections. A 10 dB shift has been applied between the maxima of the three curves for visualization purposes. (Exp: experimental data, MM: multimodal method, FEM: finite element method).

replicas has been followed. The maximal mean difference between the simulations and the experiments is 2.8 dB for the FEM and 2.2 dB for the MM.

In summary: The three configurations tested highlight the importance of taking HOM into account for high frequency simulations of the vocal tract acoustics. The effect of HOM on transfer functions and radiation patterns is emphasized by the 44CE[ɑ] and the 44EE[ɑ] configurations. The measurements and the simulations have also shown that the HOM can affect the mid-frequency behavior of vocal tract like geometries.

4.3 MRI based vowel [ɑ] geometries: influence of lips

The work presented here is a part of a study published in [54]. Measurements and FEM simulations (performed at the university La Salle of Barcelona) have been performed with two vocal tract replicas (see Figs. 3.1g and 3.1h) created from a geometry provided for the vowel [ɑ] in the Aalto database [26]. This type of geometries is among the more realistic which are used to study speech. Due to the complexity of their shape it is difficult to understand exactly how the HOM are involved in their acoustic properties. Nevertheless, they are very useful to investigate the acoustic properties of the vocal tract and to compare with the simplified geometries.

This vowel has been chosen because its large mouth aperture and its wide oral cavity allowed one to measure the pressure on a large plane inside the replicas (see Fig. 4.13). On the other hand, these geometrical features are likely to induce strong directivity effects. In order to study the influence of the lips, one replica is built without lips (see Fig. 3.1h) and the measurement and simulations obtained with both geometries are compared. A baffle of dimensions 365×360 mm was attached to the end of the replicas to mimic the face. Linear sweep signals have been used to measure the acoustic pressure at each point of a grid located in the plane (x_1, x_3) (spaced by 2.5 mm, see Fig. 4.13) from 2 kHz to 10 kHz.

4.3.1 Pressure-pressure transfer functions

In order to compare the FEM simulations with the experiments in the frequency domain, PPTF (see part 3.3.5) have been computed between several couples of points of the grid. An example of these comparisons is presented in Fig. 4.14 and Tab. 4.4. These data corresponds to the PPTF computed between a point $P_1(-0.0325, 0, 0)$ located within the oral cavity and a point $P_2(0, 0, 0)$ located at the center of the last closed cross-section which has been chosen as the origin of the coordinates. As one can see in Fig. 4.14, the experiment and the FEM simulations are quite close. The frequency and the amplitude of the three more pronounced peaks and dips of these PPTF have been extracted and are presented in Tab. 4.4. The -3 dB bandwidth of these peaks is also provided. The relative difference between the frequency of the peaks and dips f_{FEM} and f_{Exp} obtained with FEM simulations and experimentally

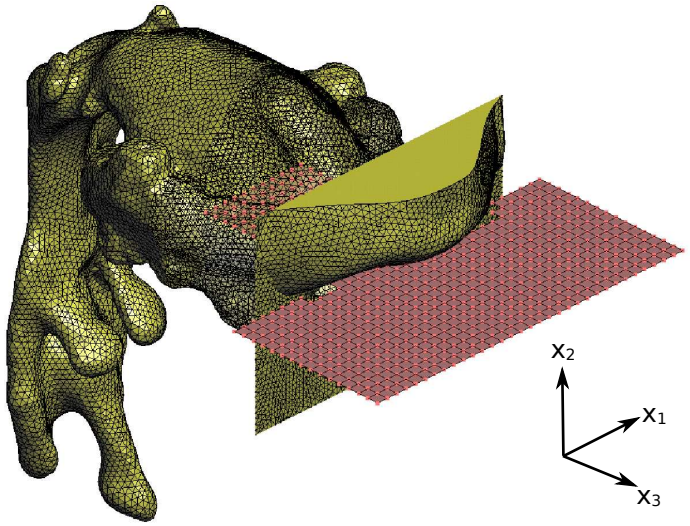


Figure 4.13: Grid of points where the acoustic pressure is measured and simulated inside a vowel [a] geometry extracted from MRI provided in the Aalto database ([26]).

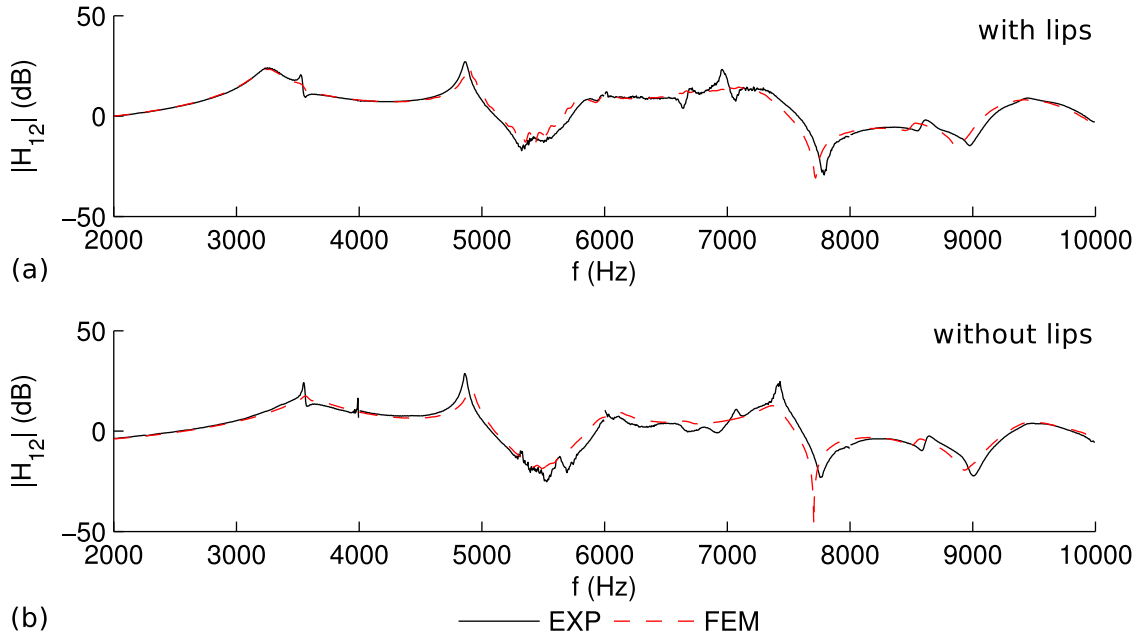


Figure 4.14: Pressure-pressure transfer function (PPTF) $H_{12}(f) = P_2(f)/P_1(f)$ for vowel [a] (a) with lips and (b) without lips obtained by finite element simulations (FEM) and experiments (Exp). $P_1(f)$ and $P_2(f)$ stand for the Fourier transform of the acoustic pressure collected at point 1 and point 3, which are, respectively, located within the oral cavity and at the mouth exit.

	With lips					
	Peaks frequencies (Hz)			Dips frequencies (Hz)		
FEM	3270	4895	8535	5335	7720	8880
Exp	3255	4865	8615	5430	7790	8980
	Peaks amplitude (dB)			Dips amplitude (dB)		
FEM	23.3	24.7	-3.5	-17.1	-30.9	-15.3
Exp	24.2	27.1	-1.9	-14.9	-43.5	-14.8
	Without lips					
	Peaks frequencies (Hz)			Dips frequencies (Hz)		
FEM	3565	4915	8575	5495	7705	8935
Exp	3550	4870	8645	5505	7760	9010
	Peaks amplitude (dB)			Dips amplitude (dB)		
FEM	17.5	20.8	-3.9	-18.6	-47	-19.4
Exp	22.6	28.3	-2.4	-28.9	-23.4	-22.5
	Peaks -3 dB bandwidth (Hz)					
	With lips			Without lips		
FEM	215	20	210	220	70	175
Exp	190 50	105	30	30	30	95

Table 4.4: Amplitude and frequencies of the three more pronounced peaks and dips of pressure-pressure transfer functions measured and simulated on two vowel [ɑ] vocal tract geometries created from MRI with and without lips between two points located respectively within the oral cavity and at the mouth exit. The -3 DB bandwidth of the peaks is also provided.

	with lips		without lips	
	FEM	Exp	FEM	Exp
3rd formant (Hz)	2205	2235	2270	2275
last formant (Hz)	9670	9755	9710	9775

Table 4.5: Frequencies of the formants selected using the procedure depicted in part 3.3.6 in order to compare the pressure field measured and simulated inside vowel [ɑ] geometries created from MRI with and without lips.

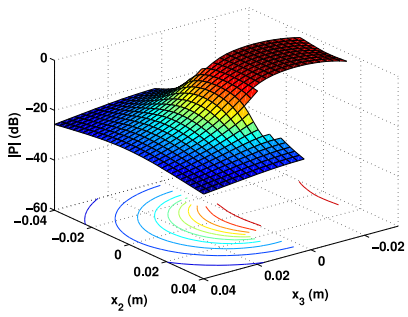
respectively, $\left(\frac{|f_{FEM}-f_{Exp}|}{f_{Exp}}\right)$, is smaller than 2%. The amplitude of the peaks tends to be higher for the experimental values (up to 7.53 dB higher for the second peak in the case without lips) and their -3 dB bandwidth tends to be smaller (down to 190 Hz smaller for the first peak of the case without lips). This could be due to the fact that the wall losses implemented in the FEM simulations are slightly over-estimated and considered as constant with the frequency. A few additional peaks and dips can be seen in the experimental data, in particular at 6640 Hz and 6975 Hz for the case with lips and 3990 Hz, 6925 Hz and 7075 Hz for the case without lips. This could be attributed to geometrical differences between the replicas and the mesh used for the FEM simulations. Indeed, though the 3D printer used is very accurate (0.025 mm to 0.05 mm per 25.4 mm), the geometry can be slightly distorted during the 3D printing process and some small amount of the wax used during the 3D printing process can remain attached to the walls of the replicas.

Except for the second dip, the frequencies of the peaks and dips are higher in the case without lips (up to 295 Hz higher for the first peak). This can be understood as the effect of the additional length introduced by the lips, which tends to lower the resonance frequencies. The dips being related to HOM propagation, the fact that the second one has a higher frequency in the case without lips could mean that the change of boundary condition induced by the presence of the lips can affect differently the HOM propagation. For the FEM simulations, the amplitudes of the peaks and dips is higher in the case with lips. With the exception of the second peak and the second dip, this is confirmed by the experiment. This can be explained by the fact that the additional length introduced by the lips induces a higher amplitude of the acoustic pressure at the mouth exit.

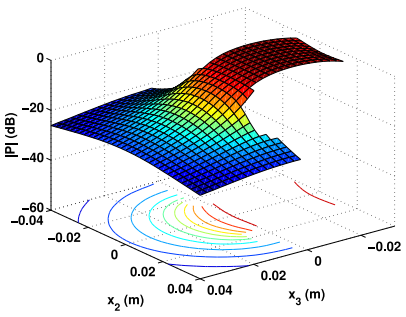
4.3.2 Pressure field maps

In order to compare spatially the FEM simulations with the experiments, pressure maps were extracted and compared. The procedure presented in part 3.3.6 has been used to choose the frequency of the comparisons. Thus, the third formant and the last formant of the frequency range have been selected (see Tab. 4.5). The corresponding pressure maps and cuts along x_2 and x_3 (in the midsagittal plane) are presented for the third formant and the last formant in Figs. 4.15 and 4.16 respectively. The cuts along x_2 were performed 1 cm in front of the mouth exit and the cuts along x_3 have been performed in the middle of the measurement surface. Iso-amplitude lines have been added to the pressure maps in order to enhance the

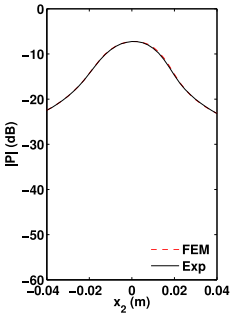
With lips



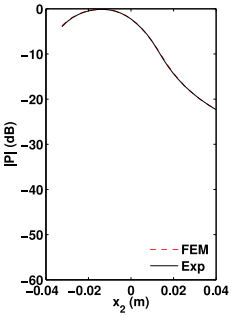
(a) FEM 2205 Hz



(b) Exp 2235 Hz

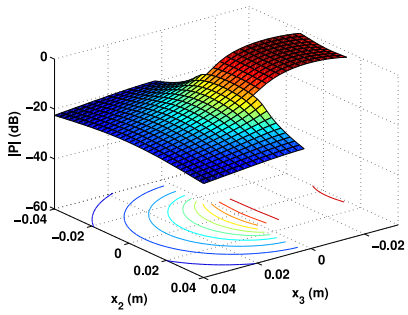


(c) Front-plane cut

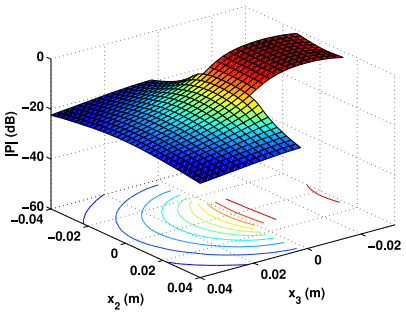


(d) Midsagittal cut

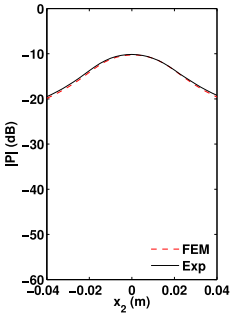
Without lips



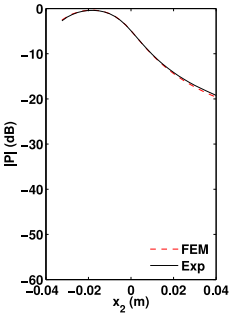
(e) FEM 2270 Hz



(f) Exp 2275 Hz



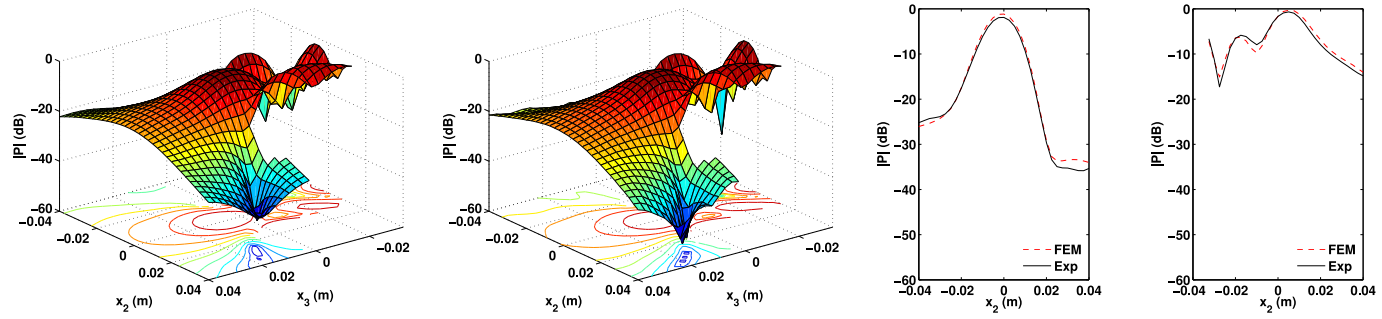
(g) Front-plane cut



(h) Midsagittal cut

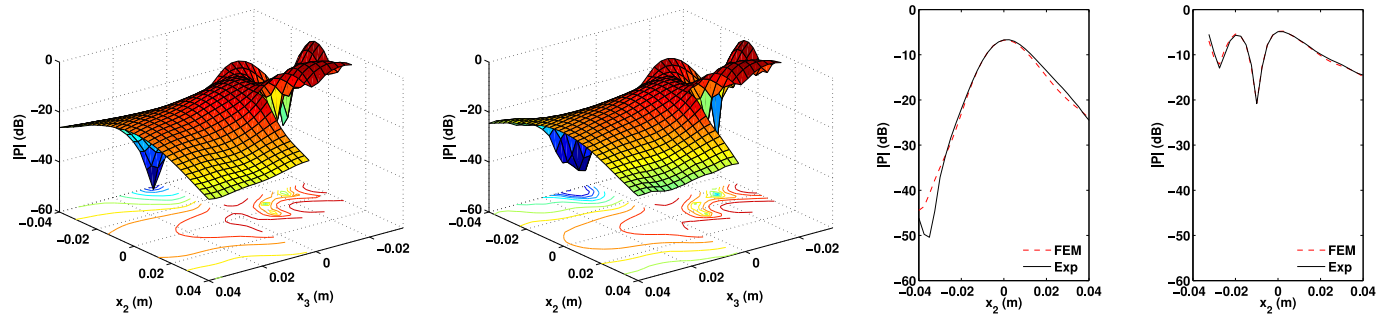
Figure 4.15: Pressure amplitude maps for the third formant of vowel [ɑ] obtained by finite element simulations (FEM) and experiments (Exp) with lips (top) and without lips (bottom). The acoustic pressure distribution within the oral cavity and at an area close to the vocal tract exit (mouth) is presented. Moreover, cuts of the pressure amplitude maps in a front-plane located 1 cm in front of the mouth exit and in the midsagittal plane are also presented.

With lips



(a) FEM 9670 Hz (b) Exp 9755 Hz (c) Front-plane cut (d) Midsagittal cut

Without lips



(e) FEM 9710 Hz (f) Exp 9775 Hz (g) Front-plane cut (h) Midsagittal cut

Figure 4.16: Pressure amplitude maps obtained for the last formant before 10 kHz of vowel [ɑ] by finite element simulations (FEM) and experiments (Exp) with lips (top) and without lips (bottom). The acoustic pressure distribution within the oral cavity and at an area close to the vocal tract exit (mouth) is presented. Moreover, cuts of the pressure amplitude maps in a front-plane located 1 cm in front of the mouth exit and in the midsagittal plane are also presented.

visualisation of the amplitude variations.

In the case of the third formant, which can be considered as low frequency, we can see that the acoustic pressure is almost uniform in the plane (x_1,x_2). Moreover, the iso-amplitude lines are perpendicular to x_3 inside the geometries. The sound is almost uniformly radiated: the iso-amplitude lines are almost like circles outside of the replica. So it can be concluded that the plane wave assumption and the baffled plane piston ([55] p226-227) are satisfying to describe the internal acoustic field and sound radiation at this frequency. There is no noticeable difference between the case with lips and the case without lips. The patterns obtained experimentally and with FEM simulations are very similar: the averaged difference between the pressure maps is of the order of 0.2 dB.

In the case of the last formant, which can be considered as high frequency, there are significant variations of the acoustic pressure in the plane (x_1,x_2). The iso-amplitude lines are no more perpendicular to the axis x_3 inside the geometries. The radiation of the sound is no more uniform: outside the geometry the iso-amplitude lines are asymmetrical with a low amplitude direction. So it can be concluded that the plane wave assumption and the baffled flat piston radiation model do not give a satisfying description of the acoustic field in this case: HOM are needed to be taken into account. In contrast to the low frequency pressure maps, significant differences of the patterns can be seen: the low amplitude direction of both configurations is located at opposite sides of the mouth exit. The radiation of these geometries is further investigated in section 5.4. Besides, there is a good agreement between the experiments and the FEM simulations: the averaged difference between the pressure maps is of the order of 1 dB.

In summary: The influence of HOM inside realistic geometries has been confirmed. The presence of the lips have been observed to lower the frequency of the peaks and dips of the PPTF, to reduce their bandwidth and to modify the radiation patterns. Eventually, the FEM successfully predicts the acoustic field inside and outside these realistic geometries.

Chapter 5

Influence of HOM on directivity

In the previous chapter the effects of the HOM have been investigated mainly inside the vocal tract. In this chapter the sound radiation outside of the vocal tract is studied. The question of how the transfer function of the vocal tract varies with the orientation of the vocal tract with respect to the position of a reception point is addressed. To do so, the radiated acoustic pressure has been measured and simulated at various angles from the exit of the vocal tract geometries with different complexity. The results presented in this chapter are partly published in [56].

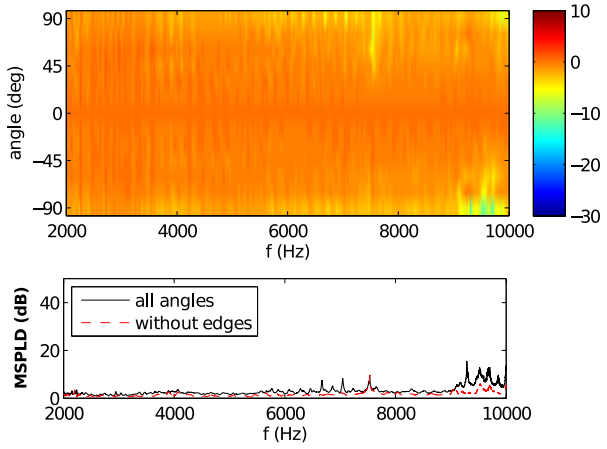
5.1 Directivity of two tubes and 44 tubes geometries with circular cross-sections

The acoustic pressure radiated by the 2CC[a], 2CE[a], 44CC[a] and the 44CE[a] has been measured at 48 cm from the exit of the replicas varying the angular position by steps of 15° using the experimental setup represented in Fig. 3.5. The distance of 48 cm was the larger distance reachable with the experimental setup and has been used as the closest to a relevant distance in the context of a conversation. All measurements have been performed in the plane (x_1,x_3), and the common edge of the 2CE[a] and the 44CE[a] has been placed on the 90° side (as represented on Fig. 3.5a).

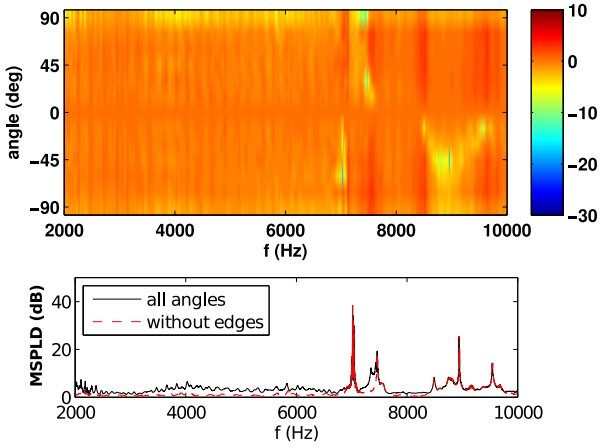
5.1.1 volume velocity distribution at the exit

In the case of the concentric configurations, the only propagation modes expected at the exit of the geometries are ψ_{00} and evanescent HOM which are generated by the discontinuity corresponding to the transition from the geometries to the exterior space. In the case of the eccentric configurations, ψ_{01} can be present and introduces the appearance of two areas with opposite phase on the exit surface.

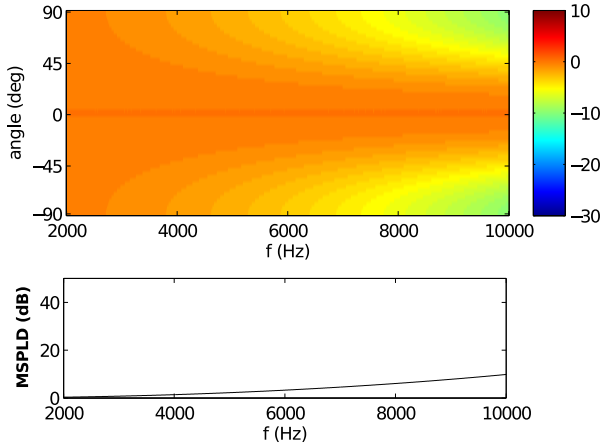
The evanescent HOM can change the volume velocity distribution on the exit surface, and one could expect an influence on the directivity of the radiated sound. However, in the concentric configurations, a simulation performed with all modes whose cut-on frequency lies below 20 kHz and a simulation with plane mode only gives exactly the same directivity patterns at 0.48 m from the exit. Thus, the influence of these modes can be neglected at this



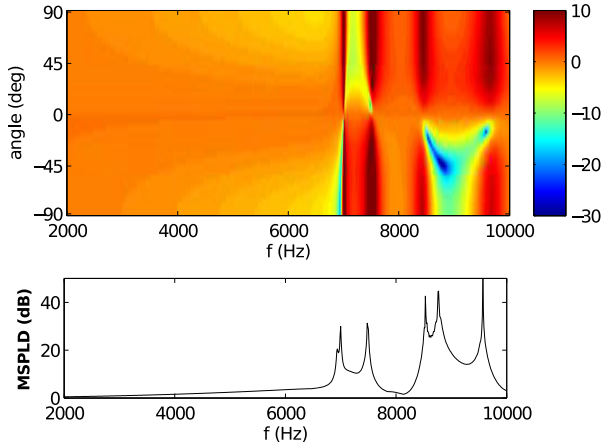
a) 2 concentric tubes (2CC[ɑ]) experiment



b) 2 eccentric tubes (2CE[ɑ]) experiment

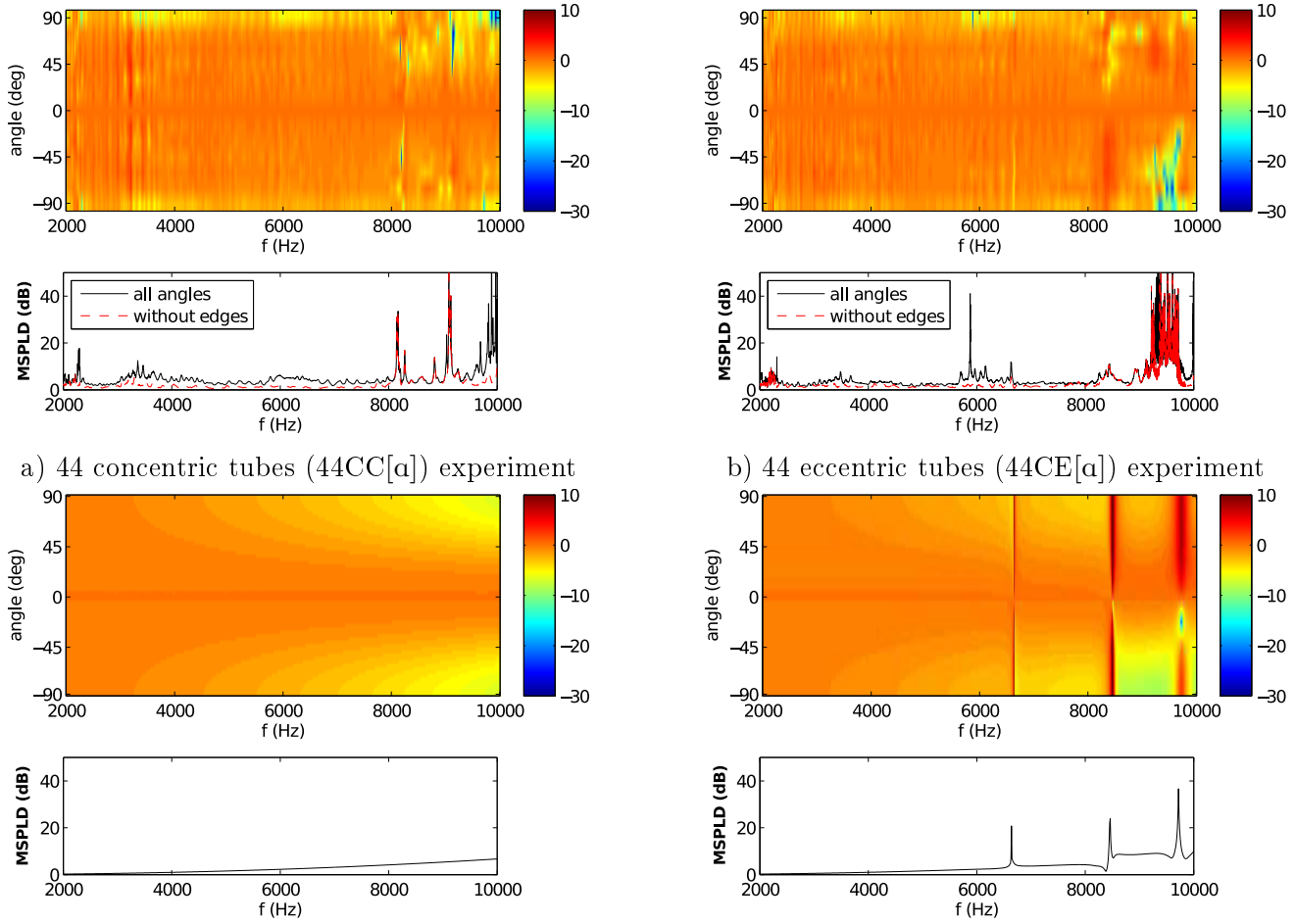


c) 2 concentric tubes (2CC[ɑ]) MM simulation



d) 2 eccentric tubes (2CE[ɑ]) MM simulation

Figure 5.1: Normalized amplitude (dB) and maximal sound pressure level difference MSPLD with respect to the angular position of the pressure radiated from two tubes vowel [ɑ] replicas with concentric (2CC[ɑ]) and eccentric (2CE[ɑ]) junction measured and simulated at 48 cm from the exit between 2 kHz and 10 kHz every 15 ° and 3 ° for respectively the experiments and the simulations. To avoid perturbations linked to experimental artifacts present on the extremal positions, the maximal sound pressure level difference has been computed with and without the edge positions (-90 ° and 90 °).



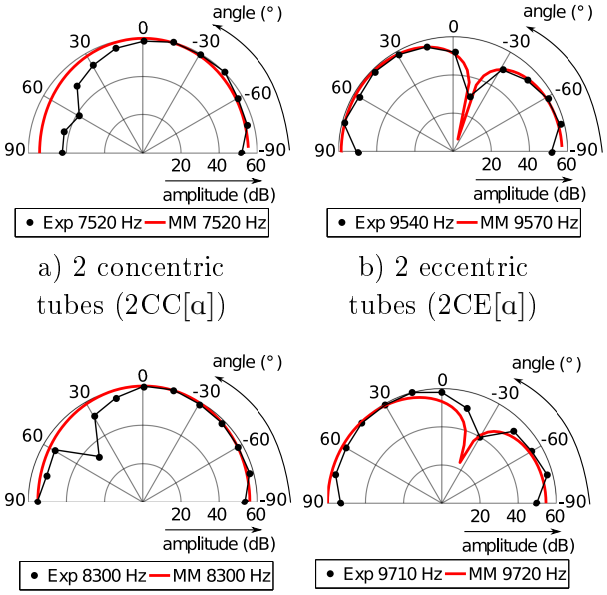
a) 44 concentric tubes (44CC[ɑ]) experiment

b) 44 eccentric tubes (44CE[ɑ]) experiment

c) 44 concentric tubes (44CC[ɑ]) MM simulation

d) 44 eccentric tubes (44CE[ɑ]) MM simulation

Figure 5.2: Normalized amplitude (dB) and maximal sound pressure level difference MSPLD with respect to the angular position of the pressure radiated from 44 tubes vowel [ɑ] replicas with concentric (44CC[ɑ]) and eccentric (44CE[ɑ]) junction measured and simulated at 48 cm from the exit between 2 kHz and 10 kHz every 15 ° and 3 ° for respectively the experiments and the simulations. To avoid perturbations linked to experimental artifacts present on the extremal positions, the maximal sound pressure level difference has been computed with and without the edge positions (-90 ° and 90 °).



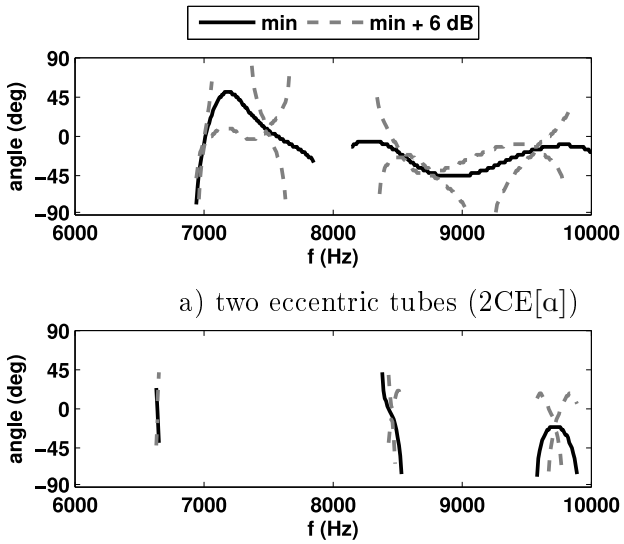
a) 2 concentric tubes (2CC[α])

b) 2 eccentric tubes (2CE[α])

c) 44 concentric tubes (44CC[α])

d) 44 eccentric tubes (44CE[α])

Figure 5.3: Measured (dotted lines) and simulated (full lines) acoustic pressure amplitude (dB) as a function of angle at selected frequencies.



a) two eccentric tubes (2CE[α])

b) 44 eccentric tubes (44CE[α])

Figure 5.4: Angle corresponding to the minimum (for maximal sound pressure level difference MSPLD values larger than 3 dB) of the radiated pressure as a function of the frequency extracted from the simulations of the eccentric geometries (2CE[α] and 44CE[α]). The angles corresponding to the pressure 6 dB higher than the minimum are presented in dashed lines.

distance and the volume velocity distribution can be considered as uniform if ψ_{00} is the only mode propagating.

5.1.2 Directivity of concentric configurations (2CC[a] and 44CC[a])

The measured and simulated normalized pressures and the MSPLD are presented in Figs. 5.1 and 5.2 for respectively the two and 44 tube geometries. Directivity patterns obtained for selected frequencies are presented in Fig. 5.3, and the angle corresponding to the minimal amplitude is presented for the simulations performed on the 2CE[a] and the 44CE[a] in Fig. 5.4.

The concentric configurations generate a one lobe symmetric directivity pattern with a maximum in the center (at 0°, see Figs. 5.1a, 5.1c, 5.2a and 5.2c). The amplitude of this lobe increases in the center and decreases towards the edges (-90° and 90°) when the frequency is increased. The MSPLD increases progressively and reaches 10 dB and 7 dB at 10 kHz for the 2CC[a] and 44CC[a] respectively. In the experimental data, this pattern is perturbed by measurement noise which can be due to insufficiently damped reflections on the walls of the insulated room, or external noise.

In addition to this noise, one can see more important localized discrepancies between the experiments and the simulations. For the 2CC[a], a peak of MSPLD of 10 dB can be seen at 7.52 kHz. The corresponding directivity pattern presented in Fig. 5.3a is asymmetric with a lower amplitude between 0°and 90°. Other less pronounced discrepancies can be found at 7 kHz and 9.3 kHz. On the experimental data obtained with the 44CC[a], peaks of MSPLD can be seen at 8.1 kHz (36 dB), 8.3 kHz (34 dB), 8.8 kHz (14 dB) and 9.1 kHz (57 dB). Inside a narrow frequency band centered on these frequencies, the directivity patterns is constituted of two asymmetric lobes separated by a low amplitude direction (see Fig. 5.3c). These discrepancies can be related to the differences observed between the measured and simulated internal pressure fields observed in sections 4.1.4 and 4.2.4 in Figs. 4.4, 4.6, 4.7 and 4.11. It can be hypothesized that they are generated by small asymmetries of the replicas, this is further investigated in section 5.2.1. Apart in these particular narrow frequency bands, the same patterns are observed for the experimental and the simulated data.

For the concentric configurations, the only propagation mode involved in the radiation is ψ_{00}. Thus, the only directivity effect which can be expected is the well known radiation from a baffled circular piston ([55] p226-227). Indeed, except for some localized discrepancies in the experimental data, the directivity pattern of these configurations is one symmetric lobe with higher amplitude in the center (at 0°) whose MSPLD increases progressively with the frequency. This pattern is only due to the difference in the distances traveled by the waves coming from the different parts of the radiating surface to the reception point which induces a phase difference.

5.1.3 Directivity of eccentric configurations (2CE[α] and 44CE[α])

For the eccentric geometries, a one lobe pattern which becomes more pronounced for increased frequency can be seen up to 6.5 kHz. Above this limit, more complex asymmetric patterns with one or two lobes are visible (see Figs. 5.1b, 5.1d, 5.2b, 5.2d, 5.3b and 5.3d). The variations of the directivity are important within short frequency intervals: the MSPLD can vary up to 50 dB within 100 Hz intervals.

In order to see better the appearance and the evolution of the two lobe pattern, the angle corresponding to the minimum between the lobes has been plotted in Fig. 5.4. This minimum is typically of the order of 30 dB lower than the maximum and is limited to a narrow angular region of about 30°. Its direction can vary a lot (up to 60°) within a short frequency interval (of the order of 100 Hz). For both 2CE[α] and 44CE[α], the minimum of amplitudes tends to be located between -90° and 0°, which corresponds to the side opposite to the common edge.

In the case of the 2CE[α], the patterns obtained by simulation can be recognized in the experimental data. For both experiment and simulation, peaks of MSPLD can be seen at 7 kHz, 7.5 kHz, 8.5 kHz, 8.8 kHz and 9.6 kHz. Except between 7.9 kHz and 8.1 kHz, two lobes can be seen at all the frequencies from 6.94 kHz (see Figs. 5.3b and 5.4a).

In the case of the 44CE[α], similar patterns can be recognized in the experiment and the simulation, but there are more differences above 9 kHz. Peaks of MSPLD can be seen at 6.6 kHz, 8.5 kHz and 9.7 kHz (see Fig. 5.3d) for both experiment and simulation. A two lobes pattern appears inside more reduced and isolated frequency intervals: 6.63 kHz to 6.65 kHz, 8.38 kHz to 8.53 kHz and 9.58 kHz to 9.89 kHz (see Fig. 5.4b). Between 6.6 kHz and 8.5 kHz a one lobe symmetric pattern can be seen and between 8.5 kHz and 9.7 kHz, there is a one lobe asymmetric pattern with lower amplitude between 0°and 90°.

In the case of the eccentric configurations, the baffled circular piston pattern can be seen under the cut-on frequency of ψ_{01} because up to this frequency ψ_{00} is the only mode propagating. Above this cut-on frequency the particle velocity distribution on the exit surface is no more uniform. The phase difference between the waves coming from the different parts of the radiating surface is not only due to the differences in traveled distance but also to the amplitude and phase variations on the radiating surface. As a consequence, more complex directivity patterns are observed.

5.1.4 Two lobes pattern

The presence of ψ_{01} at the exit of the geometries can induce the appearance of two areas with opposite phase. In this case, the radiating surface acts as an acoustic dipole and a directivity pattern with two lobes separated by a low amplitude direction is generated.

The interference of the acoustic pressure radiated by ψ_{00} and ψ_{01} can reinforce the radiated acoustic pressure on one side and decreases it on the other. Asymmetric directivity patterns are thus obtained. The low amplitude direction is no more oriented to the center (0°), but is

shifted to the side having the lowest amplitude. The variations of the amplitudes and phase of ψ_{00} and ψ_{01} with the frequency changes the direction of the minimum of amplitude.

The propagation of another HOM than ψ_{01} can generate different directivity patterns. This is what can be seen in section 5.3 in which one can see that an elliptical cross-section can generate a three lobe pattern.

5.1.5 Comparison between HOM and plane piston

At some frequencies the combination of ψ_{00} and ψ_{01} can result in a volume velocity distribution which compensates partially the phase difference due to the differences in traveled distances between the different part of the radiating surface and the reception point. This generates a reduced directivity compared to the baffled circular piston at the same frequency. This can be seen between the second and the third peaks of MSPLD and above the fifth peak of MSPLD for the 2CE[a] (see Figs. 5.1b and 5.1d). For the 44CE[a] it can be seen before or after the MSPLD peaks (see Figs. 5.2b and 5.2d). The effect of HOM is slightly noticeable below the cut-on frequency, from about 6.4 kHz, whereas the cut-on frequency of ψ_{01} is 6.8 kHz for the two tubes geometries. This can be explained by the fact that close to the cut-on frequency the reduction of amplitude of the evanescent modes generated at the junction becomes less important and they can have a significant amplitude at the exit.

5.1.6 Relation between input impedance and directivity

The peaks of MSPLD appear close to input impedance minima. For the 2CE[a] the peaks of MSPLD at 7 kHz, 7.5 kHz, 8.5 kHz and 9.6 kHz can be related to minima of impedance located at 7 kHz, 7.4 kHz, 8.6 kHz and 9.5 kHz. Likewise for the 44CE[a] the peaks of MSPLD at 6.6 kHz and 9.7 kHz can be related to minima of input impedance located at close frequencies. However, the frequencies are not exactly the same and some peaks of MSPLD cannot be related to maxima or minima of input impedance. As an example, for the 2CE[a], the peak of MSPLD at 8.8 kHz does not correspond to a maximum or minimum of input impedance. Likewise, for the 44CE[a], the peak of MSPLD at 8.5 kHz can not be related to a maximum or minimum of input impedance. One can also notice that for the 2CE[a], the peak of MSPLD located at 7.5 kHz is between a minimum (at 7.4 kHz) and a maximum (at 7.6 kHz) of input impedance. Thus, it is not possible to find a simple relation between the input impedance curve and the directivity phenomenon.

Spectrum of the sound radiated by the 2CE[a]

In order to get an idea of how the HOM effect can be perceived, the 2CE[a] replica has been excited using a broadband noise. This can be considered as an imitation of whispered voice. The spectrum of the radiated sound is presented for each measurement position in Fig. 5.5. Up to 6.5 kHz it is very similar for each position, except at the edges (-90° and 90°) where the amplitude is globally lower than at the other positions (this has the same origin as

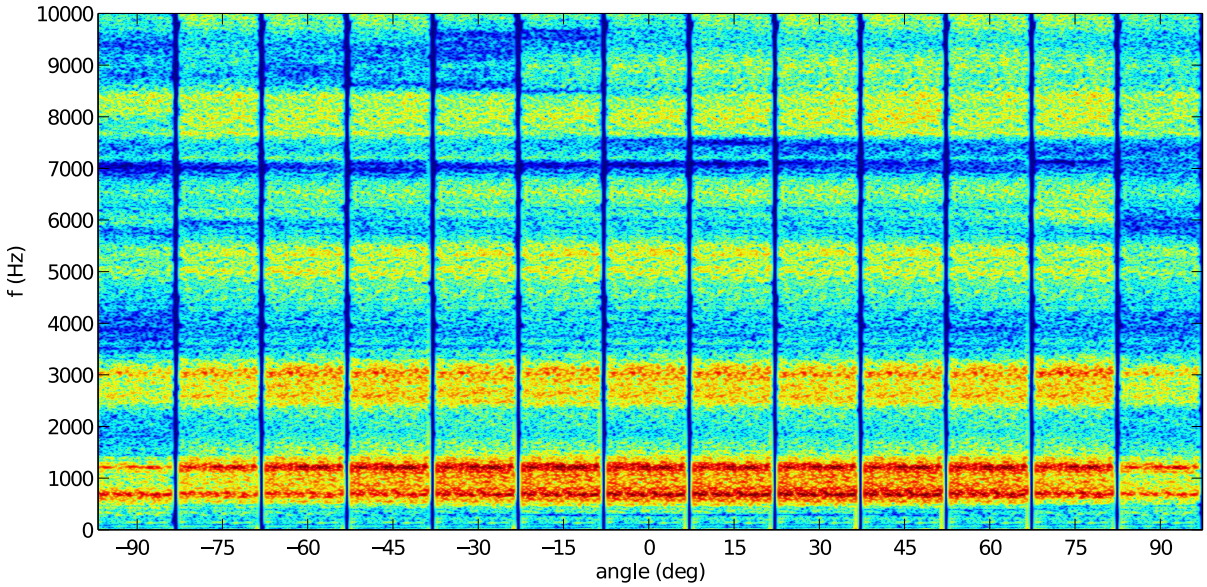


Figure 5.5: Spectrogram of the sound radiated by a vocal tract replica consisting in a two tube approximation with an eccentric junction (2CE[ɑ]) recorded at a distance of 48 cm from the exit every 15°. The excitation signal used is a broadband noise. Hann windows of 46.4 ms (2048 samples) have been used with an overlap rate of 0.9.

the artifact mentioned in section 3.3.8). Above 6.5 kHz there is noticeable differences between the positions. As an example, at -75°, -60°and -45°the amplitude is lower than at the other positions between 8.5 kHz and 9.5 kHz. The pattern observed on the directivity maps of Figs. 5.1b and 5.1d can be recognized in the variations of the spectrum with the position. The authors perceived an audible difference between the positions when listening to the recorded sounds.

5.1.7 Conclusions concerning the directivity of the 2CC[ɑ], 2CE[ɑ], 44CC[ɑ] and 44CE[ɑ]

The measurements and the simulations performed on the 2CC[ɑ], 2CE[ɑ], 44CC[ɑ] and the 44CE[ɑ] showed that the propagation of the HOM ψ_{01} inside vocal tract approximations of the vowel [ɑ] induces great variations (up to 50 dB of MSPLD) of the directivity patterns within small frequency intervals (of the order of 100 Hz) above 6.5 kHz.

A particular effect related to the mode-shape of ψ_{01} has been highlighted. The division of the exit surface into two areas with opposite phase generates a directivity pattern composed of two lobes separated by a low amplitude direction (of the order of 30 dB lower than the maximum). The variations of amplitude and phase of ψ_{00} and ψ_{01} with the frequency induce variations of the size of the lobes and the direction of the minimal amplitude (up to 60° within 100 Hz). Another effect of the propagation of ψ_{01} is the reduction of the directivity in comparison with the baffled circular piston directivity in some frequency intervals.

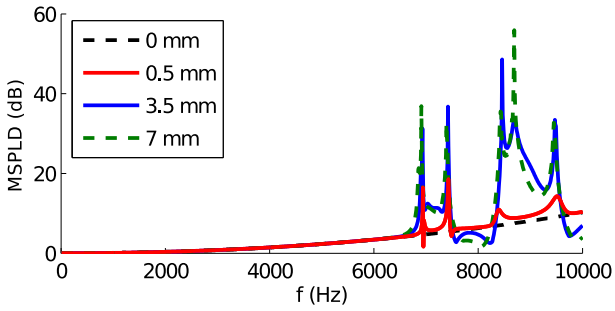


Figure 5.6: Maximal sound pressure level difference MSPLD as a function of the frequency computed for different spacing of the centers of both cross-sections of the two tubes geometries.

Since the vocal tract geometry is subject to inter- and intra-speaker differences even for a single phoneme as considered in this study, care is needed when extrapolating these conclusions to human speech. The two vocal tract approximations studied generate very different directivity patterns above the first cut-on frequency.

Because the directivity patterns become different from the baffled circular piston above the first cut-on frequency in the eccentric configurations, one can expect that the directivity is perceived differently above this frequency. This assumption is reinforced by the fact that differences related to the position are visible in the spectrogram of the sound emitted by the replicas excited with a broadband noise above 6.5 kHz. However, one can question how the great variations (up to 50 dB of MSPLD within 100 Hz) of the effects of the HOM on the directivity with the frequency are perceived or if they can be noticed by a listener. Thus, it would be interesting to carry out perceptual tests to further investigate this question. For the same reasons, classical spectral measures such as the octave/third-octave frequency band might mask the localized aspect of this effect. On the other hand, given the limited angular region of some phenomena, it would be interesting to increase the angular resolution in the future. Because of the high frequency energy present in fricative and plosive sounds, this effect is expected to have more perceptual impact. Thus, it would be interesting to investigate its consequences with vocal tract geometries corresponding to these phonemes.

5.2 Influence of eccentricity degree and convergent shape

5.2.1 Influence of the degree of eccentricity

In order to investigate the influence of the degree of eccentricity of a junction on the directivity, simulations have been performed varying the distance between the centers of the two cross-sections of the 2 tubes geometry from 0 mm to 7 mm by steps of 0.5 mm. This corresponds to the progressive transition from the 2CC[α] (Fig. 3.1a) to the 2CE[α] (Fig. 3.1b). The radiated pressure has been computed for each distance between the centers. The MSPLD corresponding to four distances between the centers is presented in Fig. 5.6. These distances, 0 mm, 0.5 mm, 3.5 mm and 7 mm, corresponds respectively to the concentric configuration, the lo-

west possible eccentricity with the discretization of the cross-sectional surface used (0.5 mm), the intermediate configuration and the fully eccentric configuration. At 0.5 mm of distance, peaks of MSPLD up to 20 dB can be seen above 6.5 kHz and the MSPLD curve is completely separated from the 0 mm curve above 8 kHz. With a distance of 3.5 mm the MSPLD curve is very similar to the 7 mm curve.

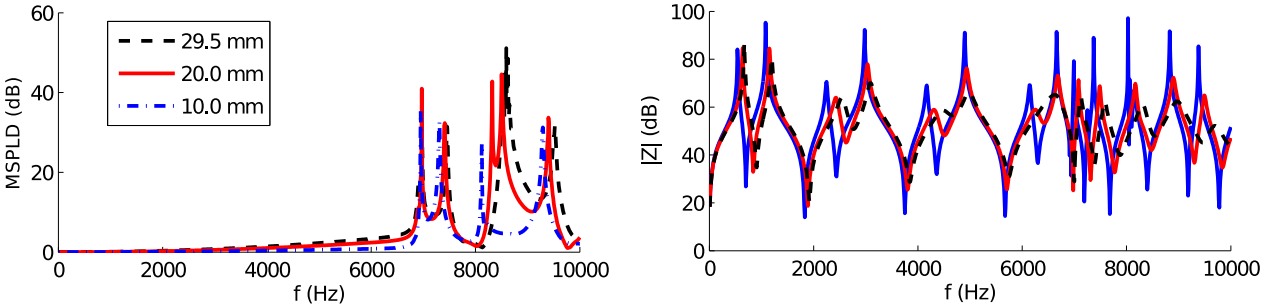
The simulations performed varying the eccentricity showed that even a very small eccentricity (0.5 mm) can induce significant changes of the directivity patterns with respect to the axisymmetric configuration (up to 15 dB of difference). Thus, it is not required to have a large eccentricity so that ψ_{01} is excited and propagate. This is in agreement with the assumption that the localized discrepancies between simulations and experiments can be explained by small asymmetries in the replicas. The fact that the frequencies of the discrepancies observed for the 2CC[a] corresponds to peaks of MSPLD of the 2CE[a] reinforces this hypothesis. However, the matching between the frequency of the discrepancies and the peaks of MSPLD of the eccentric configuration is less good for the 44 tubes geometries. The perfectly axisymmetric cases thus appear as theoretical cases difficult to reproduce accurately experimentally. The human vocal tract cannot be considered as axisymmetric and from these observations, one can conclude that axisymmetric geometrical approximations of it are not likely to reproduce a realistic directivity above their first cut-on frequency. Introducing asymmetries in such kind of approximations would allow one to simulate a qualitatively more realistic directivity above the first cut-on frequency.

In summary: The excitation of HOM being very sensitive to small asymmetries, the relevance of axisymmetric vocal tract approximation for the simulation of the speech directivity above the first cut-on frequency of the vocal tract can be questioned. Introducing asymmetries in these approximations would allow one to simulate qualitatively more realistic directivity patterns.

5.2.2 Convergent exit

During speech production, the vocal tract often has a cavity located before the mouth with larger transverse dimensions than the mouth exit, and thus has a convergent exit. This is what one can see in the area function of the 44 tubes approximations which decreases near the open end of the vocal tract (see Fig. 3.1f).

However, this is not the case for the two tubes approximation. In order to investigate how a convergent exit can affect the directivity, a third 0 length section has been added at the end of the 2CE[a] geometry. Its diameter has been reduced by steps of 0.5 mm from 29.5 mm (which corresponds to the diameter of the second section) to 10 mm. A convergent exit is thus gradually introduced. The radiated pressure has been computed for each diameter value. The MSPLD and the input impedance corresponding to 3 diameters values, 29.5 mm, 20 mm and 10 mm, are presented in Figs. 5.7a and 5.7b respectively. The first one corresponds to the 2CE[a] without convergent exit.



a) Maximal sound pressure level difference b) Input impedance

Figure 5.7: Maximal sound pressure level difference MSPLD and input impedance as a function of the frequency computed for different diameters of the exit of the two eccentric tubes geometry.

When the diameter of the third section is reduced, the MSPLD is globally decreasing. With a diameter of 10 mm it is almost zero up to 6.5 kHz. However, above 6.5 kHz the MSPLD still has locally important values (up to about 35 dB at 7 kHz, 7.3 kHz, 8.1 kHz and 9.3 kHz for an exit diameter of 10 mm), but between these peaks the MSPLD is reduced. The frequency of the peaks is decreased and one peak disappears at 10 mm. The peaks of input impedance have a larger amplitude and a reduced bandwidth and the minima have a reduced amplitude. The frequency of the peaks and minima of input impedance is also reduced.

In the case of a convergent exit, the radiation losses are reduced in comparison with a non-convergent exit. As it can be observed on the input impedance curves, this induces a higher amplitude and a reduced bandwidth of the resonances inside the vocal tract. This can be related to the fact that the effects of the HOM are more localized around the peaks of MSPLD for reduced exit diameters (see Fig. 5.7). When the diameter of the exit is reduced, one gets closer to a closed end boundary condition and the resonance frequencies are reduced. This can be related to the decrease of the frequency of the peaks of MSPLD and input impedance for reduced exit diameters. Thus, the fact that the 44CE[a] has a convergent exit and not the 2CE[a] can explain why the effects of HOM are more localized in frequency for the 44 eccentric tubes geometry.

The influence of HOM can be seen even with a diameter of 10 mm which is almost three times smaller than the diameter of the large tube. Thus, more generally, it can be expected to be significant even if the mouth aperture is small compared to the preceding cavity. One can expect that the influence of HOM on the directivity varies qualitatively with the phonemes. A wide mouth aperture would induce an effect visible at all the frequencies above the first cut-on frequency of the vocal tract and a narrow mouth aperture would induce an effect limited to localized narrow frequency intervals. This is further investigated in section 5.5.

In summary: The effects of HOM can be seen even if the radiating surface is small compared to the largest dimension of the vocal tract. However, in this case, they tend to be localized in particular frequency bands.

5.3 Directivity of a configuration with elliptical cross-section (44EE[α])

Directivity measurements have also been performed on the 44EE[α], however, in this case, in order to better observe the phenomena depicted in section 5.1, the angular resolution has been increased to 3°. To do so, the experimental setup using the positioner has been used (Fig. 3.4) instead of the one using a manually moved microphone. The motion of the probe-microphone being automated, the measurement of the acoustic pressure at more locations was less tedious and more accurate than by moving the microphone manually. However, in this case, the maximal achievable distance from the exit is limited to 10 cm. As a consequence, the directivity can be measured only in relatively close field. In order to investigate the influence of the HOM on directivity at higher frequencies than 10 kHz, the frequency range has been increased to 15 kHz.

5.3.1 Comparison with experiment

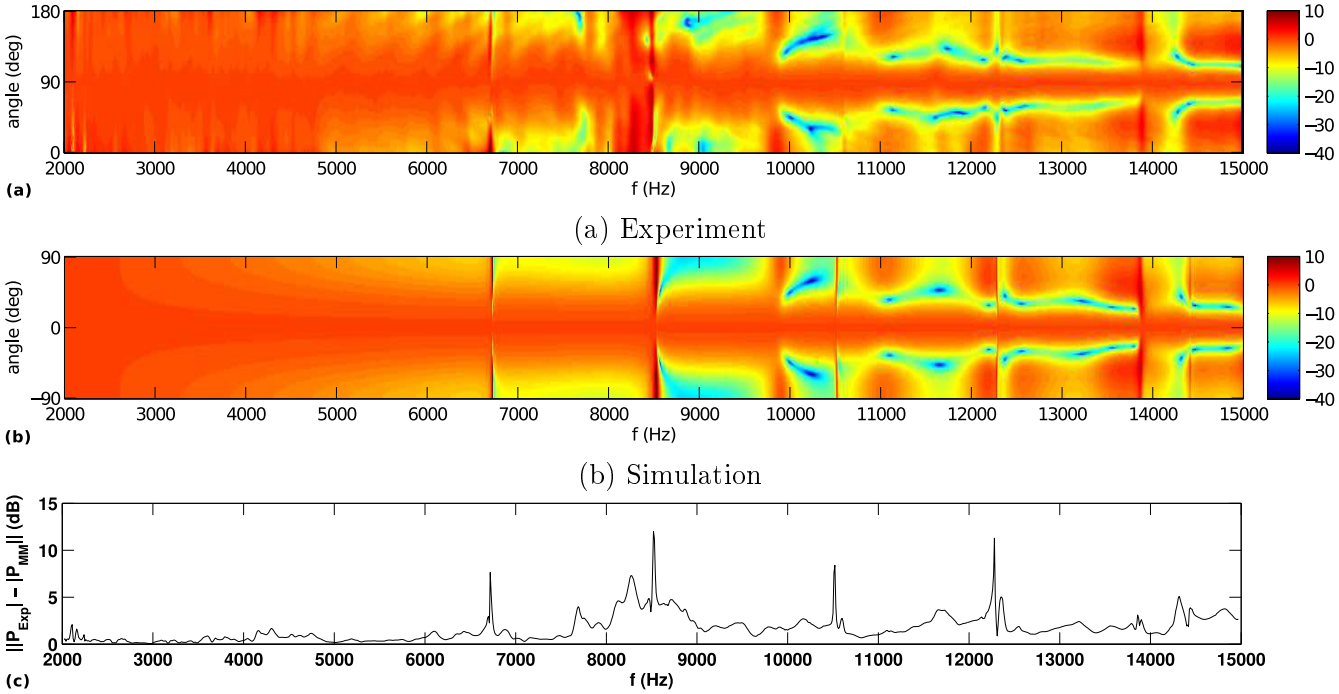
The directivity maps measured and simulated with MM in the horizontal plane (x_1,x_3) on the 44EE[α] at 4 cm from the exit as well as the averaged difference between the experiment and the simulation are presented in Fig. 5.8.

Below 6.7 kHz, the effect of the diffraction of the plane waves can be seen as a one symmetric lobe pattern which becomes more pronounced when the frequency is increased.

Above the cut-on frequency of ψ_{e21} (6084 Hz at 22.65°C, see Fig. 2.4), the directivity patterns become more complex. As for the 2CE[α] and the 44CE[α], they show important change within small frequency intervals (of the order of 100 Hz) and small angular regions (of the order of 30°). In particular, transitions from almost omnidirectional to very directional patterns within less than 100 Hz are observed at 6.72 kHz, 8.49 kHz, 10.61 kHz, 12.31 kHz and 13.92 kHz for the experiment and at 6.72 kHz, 8.52 kHz, 10.52 kHz, 12.29 kHz, 13.9 kHz and 14.41 kHz for the simulation. However, unlike the geometries with circular cross-sections, three lobe patterns are observed instead of two lobe ones.

Unlike the geometries with circular cross-section, the three lobe pattern can be obtained with the plane mode only from 9.4 kHz. Indeed, the wavelength (3.67 cm at 9.4 kHz) is smaller than the width of the exit of the geometry (4.42 cm) in this frequency range. And the phase variations along the width of the exit are larger than 2π. However, this three lobe pattern is perturbed by the propagation HOM. The position of the amplitude minima is modified, and the amplitude of the lobes can be amplified or reduced depending on the frequency. The one lobe pattern is similarly perturbed by the propagation of HOM between 6.7 kHz and 9.4 kHz.

Globally, the directivity maps obtained by experiment and simulations show very similar patterns which are perfectly symmetric with respect to the 0° position for the simulations and with small asymmetries for the experiment. The amplitude maxima and minima have similar angular position and occur at similar frequencies. This good agreement, is confirmed

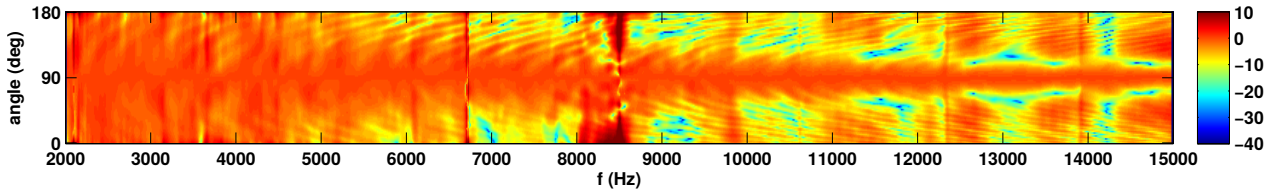


(a) Experiment

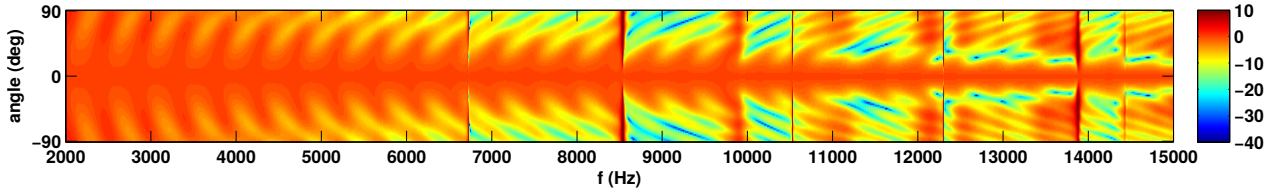
(b) Simulation

(c) Difference between the experiment and the simulation averaged on the angular position

Figure 5.8: Normalized amplitude as a function of the frequency and the angular position measured and simulated at 4 cm from the exit of a vowel [ɑ] vocal tract replica created from an area function from [23] constituted of 44 tubes with elliptical cross-sections with junction eccentric in the vertical plane (x_2,x_3) and symmetric in the horizontal plane (x_1,x_3) (44EE[ɑ]. The difference between the experiment and the simulation averaged on the angular positions is also presented as a function of the frequency.



(a) Experiment



(b) MM simulation with reflection

Figure 5.9: Normalized amplitude as a function of the frequency and the angular position measured and simulated at 10 cm from the exit of a vowel [ɑ] vocal tract replica created from an area function from [23] constituted of 44 tubes with elliptical cross-sections with junction eccentric in the vertical plane (x_2,x_3) and symmetric in the horizontal plane (x_1,x_3) (44EE[ɑ]). The simulation takes into account a reflection on a plane located 37.6 cm behind the microphone.

by the fact that the difference between the experiment and the simulation averaged on the angular position (see Fig. 5.8c) is globally smaller than 5 dB. There are however some peaks of averaged difference with higher amplitude (at 6.72 kHz, 8.28 kHz, 8.52 kHz, 10.52 kHz and 12.28 kHz). But they are due to small differences in the frequency at which abrupt changes in directivity pattern occur. The averaged difference globally increases with the frequency. This can be related to the fact that the more complex patterns do not match exactly: there are small differences in the angular position of the minima of the three lobe pattern (of the order of 5°).

The small asymmetries observed in the experiments can be attributed to small imperfections of the replica, insufficiently damped reflections or background noise. The difference in the frequency at which the abrupt changes of directivity pattern occur can be explained by a difference between the cut-on frequency of the HOM used for the simulations and the experimental one. This can come from inaccuracies in the measurement of the temperature, variations of the temperature during the experiment (the cut-on frequency of the HOM is very sensitive to temperature change[1]), difference between the experimental boundary condition and the one implemented in MM (the 3D printed material can be slightly porous) and from numerical errors. These small frequency differences explain the presence of peaks in the averaged difference curve at the same frequencies.

5.3.2 Effect of first reflection outside the replica

The acoustic pressure has been measured and simulated in the horizontal plane (x_1,x_2) on the 44EE[α] at a larger distance from the exit of 10 cm. The corresponding directivity map is presented in Fig. 5.9.

Directivity patterns globally similar to the ones of Fig. 5.8 are observed: abrupt transitions within small frequency intervals (order of 100 Hz) and three lobe patterns. However, in this case a periodic pattern, shaped like parabolic arcs is visible on the measured directivity maps on all the frequency range (see Fig. 5.9a).

It has been hypothesised that this is the result of the interference with reflected waves. Thus, a reflection have been introduced in the simulation following the method described in section 2.3.2.2 in order to reproduce this phenomenon. Since the arm of the positioner has been estimated as the object which most likely causes reflections, the reflection plane of the simulation has been placed at its location which is 37.6 cm behind the end of the probe-microphone. Note that the displacement of the reflection plane with the motion of the positioner is taken into account in the simulation.

On the simulated directivity map (see Fig. 5.9b) one can see the directivity patterns with a periodic interference pattern shaped like parabolic arcs superimposed. Thus, even though the interference patterns obtain by simulation are not exactly similar to the one observed on the experimental data, one can conclude that the phenomenon is qualitatively reproduced and that the hypothesis of the interference with a reflected wave is consistent.

The differences between the interference patterns of the experiment and the simulations can be attributed to the limitation of the simple specular model used for this modelization, to the fact that the reflection coefficient of the surface is not one, the surface of the positioner harm is not rigorously flat and that other reflections not taken into account also inteferes with the direct field. Indeed, a more careful observation of Fig. 5.9a reveals that a second periodic parabolic pattern is superimposed to the more important one. The arcs of this second pattern are narrower and closer to each other. Since it has been observed with MM simulations that the parabolic patterns becomes narrower and closer to each other when the reflection plane is farther from the microphone, it can be hypothesised that this pattern is due to a reflection on the wall located in front of the replica.

The influence of the reflections is more important and more visible at 10 cm from the exit than at 4 cm because the direct field has a lower amplitude at this distance, whereas the reflected waves can have, on the contrary, a higher amplitude. However, a more careful observation of Fig. 5.8a reveals that an interference pattern is also visible between 5 kHz and 8 kHz. The variations of amplitude of the interference pattern related to the reflections can reach 25 dB at 10 cm, whereas it is limited to 5 dB at 4 cm.

However, this effect remains lower in amplitude than the diffraction of the plane waves and

[1] A variation of temperature of 1°C induces a variation of about 10 Hz of the frequency of ψ_{01} in the case of a circular cross-section of diameter 29.5 mm.

the influence of HOM. Indeed, without reflections the MM simulations can have a maximal difference up to 50 dB. On the other hand, the measurements performed at 48 cm (see section 5.1) show that even though there are non negligible reflection at this distance, the same patterns are observed on the directivity maps measured and simulated. The fact that no interference patterns are visible on the directivity maps of Figs. 5.1 and 5.2 can be explained by the lower angular resolution, which does not allows one to observe it. However, there are some small periodic amplitude variations for each position which may be explained by reflections.

The fact that even in an environment designed so that the reflections are as limited as possible interference patterns due to reflection significantly influence the observed directivity maps shows that this phenomenon cannot be neglected when considering a more realistic context in which nothing is done to damp the reflections. Eventually, this raises the question of how to account for the room acoustics with the radiation model used.

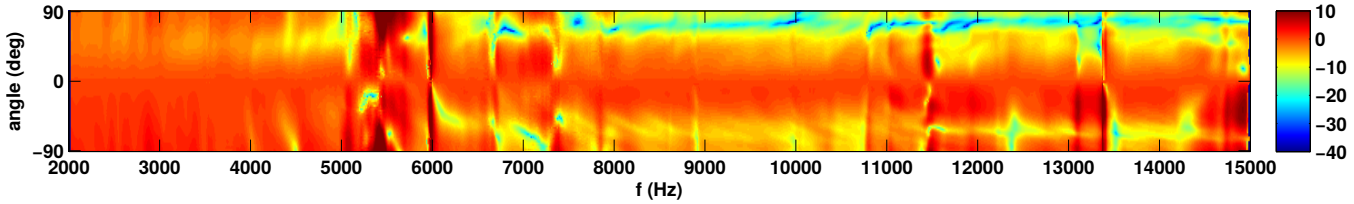
5.4 Directivity of the MRI based vocal tract geometries

In order to investigate how the sound is radiated from the vocal tract with geometries as realistic as possible, directivity measurements have been performed on the geometries created from MRI (see Figs. 3.1g and 3.1h). As for the 44EE[a], the setup using a positioner has been used in order to insure a good angular resolution (3°) and the measurements have been performed in the frequency range from 2 kHz up to 15 kHz.

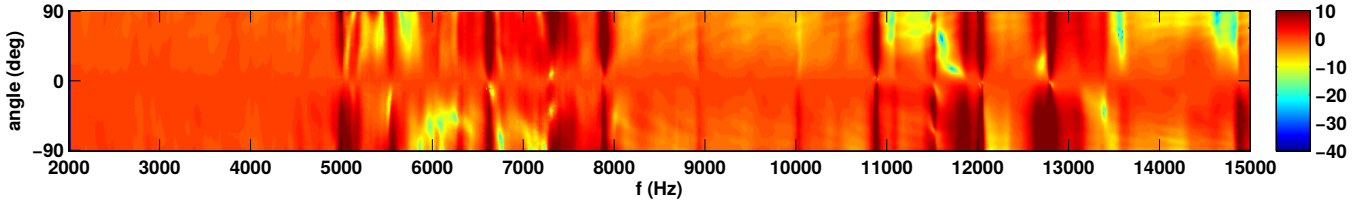
5.4.1 Directivity measurements performed with realistic replicas

The amplitude of the sound radiated by both MRI based geometries is presented as directivity maps in Fig. 5.10.

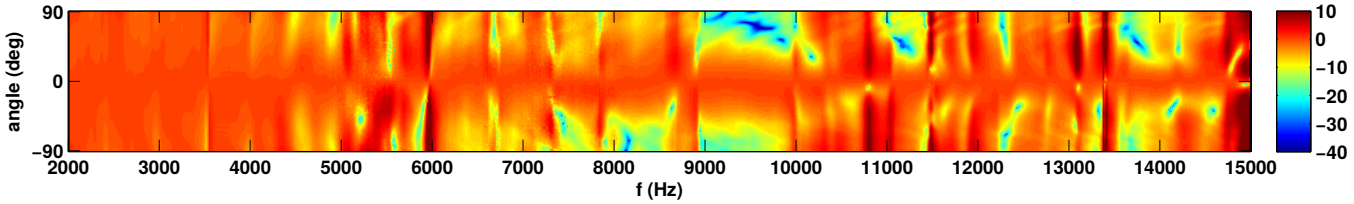
Except in the case of the vertical plane (x_2,x_3), the directivity patterns are almost omnidirectional up to 3.5 kHz. The first effect which can be attributed to HOM propagation is visible in the horizontal plane at 3.56 kHz for the case with lips and 3.58 kHz for the case without lips. It appears as an abrupt change (within less than 20 Hz) from a one lobe pattern with lower amplitude on the -90° side to a lower amplitude on the 90° side. The effects of HOM can then be seen at all the frequencies from about 4 kHz for the case with lips and about 5 kHz for the case without lips. As observed in sections 5.1 and 5.3, it appears as significant variations of the directivity pattern within short frequency intervals (of the order of 100 Hz) and small angular regions (of the order of 30°). However, in this case, it can be seen in larger proportion and complexity than with the 2CE[a], the 44CE[a] and the 44EE[a]. Indeed, more than ten abrupt transitions of directivity pattern can be seen for both geometries in both planes, in particular at 5.96 kHz, 11.49 kHz and 13.37 kHz for the lips case and at 6.63 kHz, 10.89 kHz and 12.8 kHz for the no lips case. Directivity patterns with one, two or three lobes are observed. They are asymmetric at most of the frequencies.



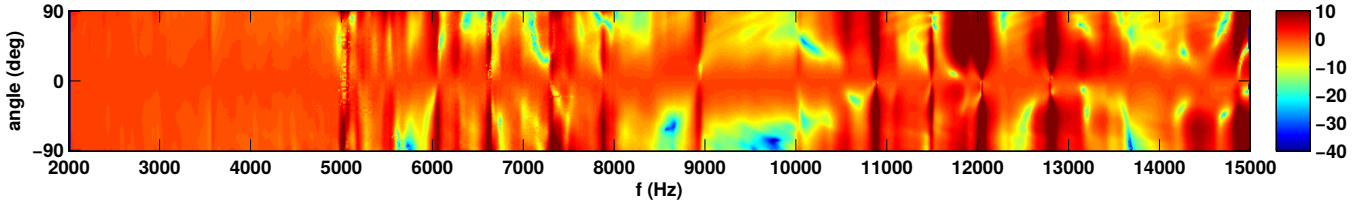
(a) With lips vertical (x_2,x_3)



(b) Without lips vertical (x_2,x_3)



(c) With Lips horizontal (x_1,x_3)



(c) Without Lips horizontal (x_1,x_3)

Figure 5.10: Normalized amplitude (dB) as a function of the frequency and the angular position of the acoustic pressure radiated at 4cm from the exit of two replicas built from an MRI database ([26]) with and without lips (see Figs. 3.1g and 3.1h). These measurements have been performed in the horizontal (x_1,x_3) and the vertical (x_2,x_3) planes.

The greater complexity of the measured directivity patterns compared to the ones observed with the 2CE[ɑ], the 44CE[ɑ] and the 44EE[ɑ] can be explained by the greater complexity of the MRI geometries. The fact that almost no symmetric patterns are observed shows that the symmetric geometries are particular cases which does not match the behaviour of a realistic vocal tract shape. The lower frequency limit at which the HOM effect starts compared to the previously studied replicas can be explained by the wider dimensions and the overall asymmetry and eccentricity of the shape which allows every propagation mode, even the ones with the lowest cut-on frequency, to be excited and to propagate.

5.4.2 Effect of the lips on the directivity

Significant differences can be observed between the case with lips and the case without lips. The number and the frequency of the abrupt change of directivity pattern are different. However, some similar patterns can be found at close frequencies, as an example in the horizontal plane (x_1,x_3) at 3.56 kHz for the case with lips and at 3.58 kHz for the case without lips. A similar global tendency can be observed in the horizontal plane. See as an example the area with no abrupt change of directivity pattern between 9 kHz and 10 kHz. However, in this case the minimum is not located on the same side (90° in the case with lips and -90° in the case without lips).

There are more differences in the vertical plane (x_2,x_3). In the case with lips two directions with globally lower amplitude (at about -70° and 70°) which are not present in the case without lips are observed. The amplitude also tends to be globally higher in the center. This can be understood as the effect of the lips, which is more important in the vertical plane (x_2,x_3) because it is in this plane that the presence of the lips induces the more important geometrical change. One can also notice that, in the case with lips, the 90° side has globally lower amplitude at all the frequencies. This can be explained by the fact that the upper lip, which is located at the 90° side, is larger than the lower lip in this geometry. The fact that the lower amplitude induced by the presence of the upper lip can be seen even at the lowest frequency shows that the diffraction by the upper lip affects the directivity even if the only propagation mode present is the plane mode.

In summary: The lips appear as a key element in the radiation of the sound, especially in the vertical plane (x_2,x_3). The combination with the effect of the HOM coming from inside change significantly the radiation patterns. However, this also introduces some global changes which can be seen at all the frequencies.

5.5 Comparison of the vowels [ɑ], [i] and [u]

After having investigated the effect of the HOM on various degrees of simplification of the vowel [ɑ] this vowel has been compared to two other vowels, [i] and [u]. These three vowels

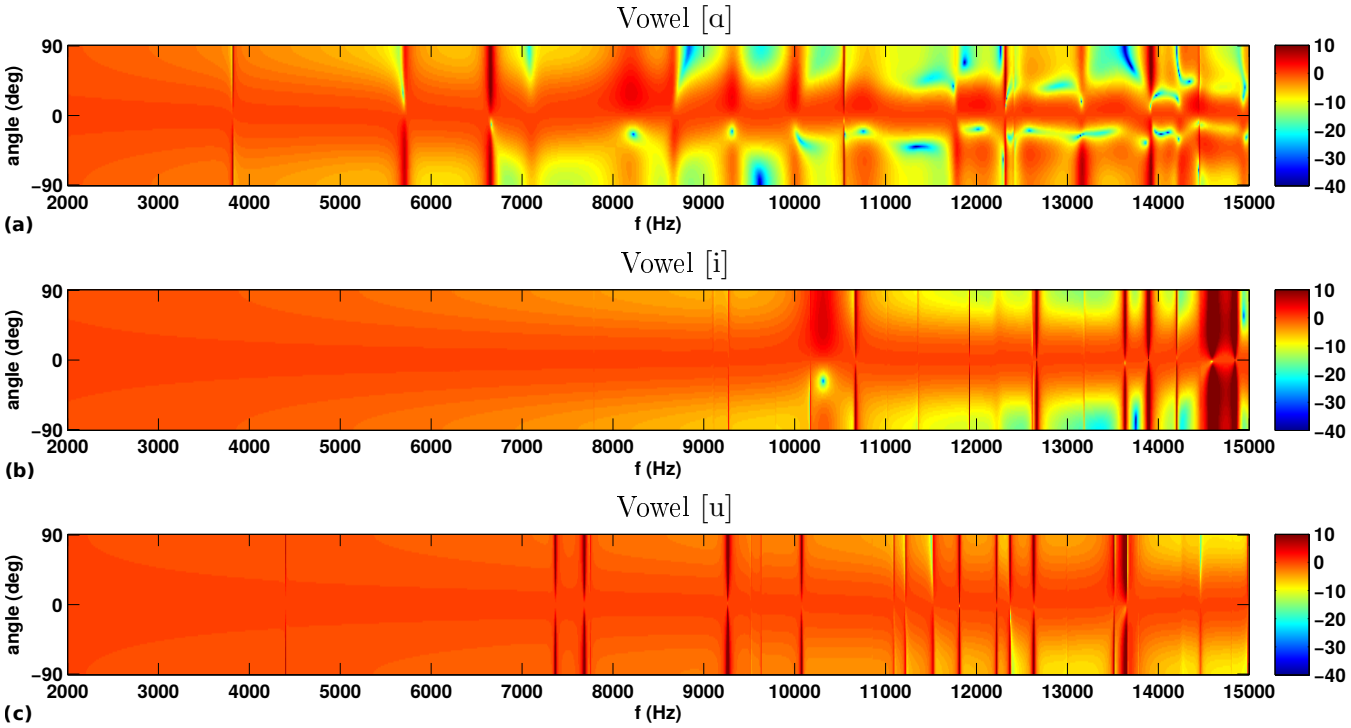


Figure 5.11: Normalized amplitude (dB) as a function of the frequency and the angular position simulated at 60 cm from the exit of three vocal tract approximations for the vowels [ɑ], [i] and [u] (44EFE[ɑ], 44EFE[i] and 44EFE[u]) created from area functions provided by Story [23].

corresponds to the three extremes of the possible geometrical shapes of the vocal tract. As for the 44EFE[ɑ], the 44EFE[i] and 44EFE[u] have been created from area functions provided by Story [23]. The conclusions drawn from the preceding results lead one to create it with elliptical cross-sections and fully eccentric junctions in the plane (x_2,x_3) and 25 % eccentric junctions in the plane (x_1,x_3) (see Fig. 3.3). The 25 % eccentricity in the plane (x_1,x_3) has been added in order to break any symmetry in the geometries.

5.5.1 Simulation of the directivity patterns of the vowels [ɑ], [i] and [u]

The acoustic pressure has been simulated with the MM at 60 cm from the exit of the geometries up to 15 kHz for a temperature of 20°C. The corresponding directivity maps are presented in Fig. 5.11.

As for the previously studied geometries, one can observe the combination of the effect of the diffraction of the plane mode and the effect of the HOM. For the three geometries, abrupt transitions occurring in a small frequency interval (of the order of 100 Hz) are observed. The frequency intervals in which the abrupt transitions occur are narrower in the case of the 44EFE[u] and even narrower in the case of the 44EFE[i]. The effect of HOM begins to be

visible at 3.8 kHz for the 44EFE[ɑ], 9.1 kHz for the 44EFE[i] and 4.4 kHz for the 44EFE[u]. Their effect tends to be localized in narrow frequency intervals in the case of the 44EFE[i] and the 44EFE[u]. In the case of the 44EFE[ɑ], the effects of HOM are present at all the frequencies from 6.7 kHz and inside two narrow intervals (of the order of 500 Hz) located around 3.8 kHz and 5.7 kHz. In the case of the 44EFE[i], the effect is present inside two very small frequency intervals (about 50 Hz) above 9.1 kHz, and inside wider intervals above 10 kHz. There is a strong minimum located around 10.3 kHz in the direction -26°. In the case of the 44EFE[u], the effect is limited to 20 narrow bands (smaller than 300 Hz).

5.5.2 Comparison between the approximations of the vowel [ɑ] and the measurements performed on the realistic replicas of the vowel [ɑ]

The directivity map simulated with the 44EFE[ɑ] is qualitatively closer to the directivity maps measured on the MRI based replicas than the other vowel [ɑ] approximations. Indeed, in the case of the 44EE[ɑ] the directivity patterns show complex variations with abrupt changes, but it is perfectly symmetric with respect to the central position 0° (see Figs. 5.8a and 5.8b). On the opposite, the directivity map of Fig. 5.11a shows asymmetric directivity patterns for all the frequency ranges in which the effect of HOM is observed. Another similarity between the 44EFE[ɑ] and the realistic geometries is the fact that the first effect of HOM is observed at 3.8 kHz and 3.5 kHz and that it is observed at all the frequencies from 6.7 kHz and 5 kHz. On the opposite, the first observed effect of HOM for the 44EE[ɑ] is at 6.7 kHz and the effect is visible at all the frequencies from 8.5 kHz.

The same qualitative behavior is observed for the 44EFE[ɑ] and the MRI based replicas with more complexity for second one. Thus, one can conclude that the use of elliptical cross-sections with eccentricity in both planes allows one to simulate acoustic properties qualitatively close to the one of realistic geometries using a very simple description of the vocal tract shape based on an area function. This can be of particular interest if one desire to perform simulations with a lower computational cost than with numerical methods such as FEM or FD. The simple description of the geometry is also an advantage: an area function is a very small amount of data compared to a 3D geometry based on MRI and it can be interpolated. Simple articulatory models could also be used for this purpose.

5.5.3 Comparison between the vowels [ɑ], [i] and [u]

The 44EFE[ɑ] has a mouth opening wider (4.42 cm) than the one of the 44EFE[i] and the 44EFE[u] (2.51 cm and 2.4 cm respectively), which is preceded by a large oral cavity (6.2 cm in the widest part, see Figs. 3.2 and 3.3a). Thus, HOM are expected to propagate at relatively low frequency (below 4 kHz) and they can influence directly the volume velocity distribution at the exit of the geometry. As a consequence, more complex radiation patterns with a more pronounced directivity are generated. In addition, the larger distance between the different radiating points of the surface induces a more pronounced directivity due to the diffraction of the plane waves, which increases the overall directivity.

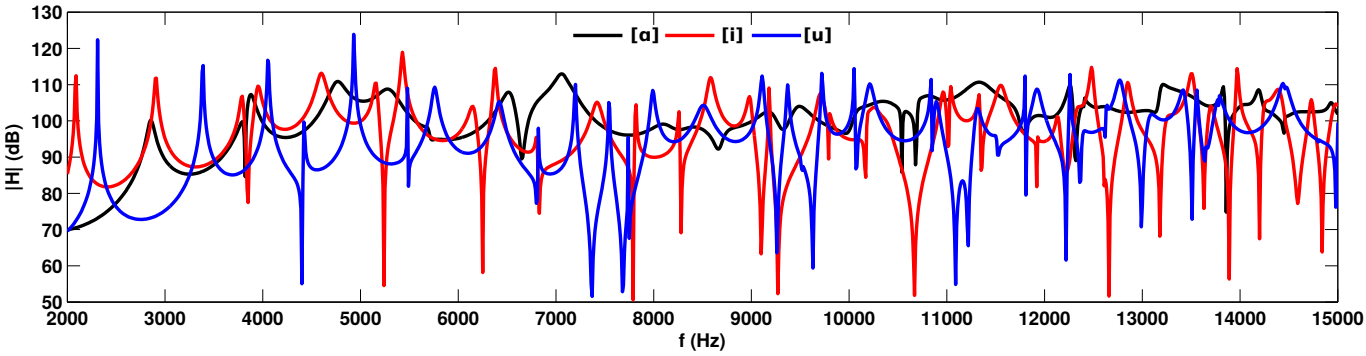


Figure 5.12: Transfer functions between the input volume velocity and the acoustic pressure radiated at 60 cm in front of the exit of three vocal tract geometries corresponding to the vowels [ɑ], [i] and [u] (44EFE[ɑ], 44EFE[i] and 44EFE[u]) created from area functions provided by [23].

In the case of the 44EFE[u], there is a large cavity (5.67 cm in the widest part) preceding a small mouth opening (2.4 cm) (see Figs. 3.2 and 3.3c). This allows the HOM to propagate at relatively low frequency (4.4 kHz), but their influence on the radiation is limited to small frequency intervals (of the order of 500 Hz) because of the small opening. The effect of a convergent exit investigated in section 5.2.2 is thus observed and confirmed here.

In the case of the 44EFE[i], there is a cavity with large dimensions (6 cm in the widest part) near the entrance of the geometry (between $x_3 = 2$ cm and $x_3 = 10$ cm) which is connected to the mouth by a narrow channel. The HOM which can propagate at relatively low frequency in the large cavity can influence the acoustic field at the entrance of the channel and excite evanescent HOM, but they are exponentially damped along this channel and until a relatively high frequency (9.1 kHz) only plane waves arrive at the opening. The effect of HOM begins to be visible only for the frequencies at which the wavelength is of the same order of size as the width of the channel (of the order of 2 cm). The effect of the HOM propagating inside the wider part can then be transmitted to the opening. The volume velocity distribution at the exit is thus influenced by both the HOM of the widest part which induces abrupt changes in narrow frequency intervals and the HOM propagating in the channel which induce effects more spread in the frequencies.

In summary: The effect of a convergent exit investigated in section 5.2.2 is confirmed here. One can see that it is possible to predict qualitatively the effect of HOM on the directivity of radiated sound from the shape of the vocal tract geometry.

5.5.4 Transfer function variations

In order to investigate the effect of HOM on the transfer functions of the three vowels, the transfer functions between the input volume velocity and the acoustic pressure radiated at 60 cm in front of the geometries (at 0°) have been computed. They are presented in Fig. 5.12.

One can notice the presence of anti-resonances due to HOM propagation at relatively low frequency for the three geometries. In the case of the 44EFE[a] and the 44EFE[i], the first anti-resonances occur at 3.85 kHz and in the case of the 44EFE[u] it occurs at 4.4 kHz. Other anti-resonances and resonances due to HOM propagation can be found above these frequencies. Thus, even if the propagation of HOM does not necessarily impact the directivity of the radiated sound, effects on the transfer functions are present in all the cases.

However, one can not consider a unique vocal tract transfer function when the directivity effects are significant. When the directivity is only produced by the diffraction of the plane waves, the one lobe pattern generated does not induce significant variations of the transfer function. The effect produced can be modeled by a low pass filter whose cut-off frequency reduces towards the edges (-90° and 90°). When HOM can affect the directivity, the effect produced on the variations of the transfer function with respect to the location of the reception point are more complex. At a given frequency, the transfer function can have a minimum at one angular position and a maximum at the other. It is no more possible to define a unique transfer function which can be modified to account for plane wave diffraction. There are as many vocal tract transfer functions as reception points.

In summary: Depending on the shape of the vocal tract, its transfer function can have more or less important variations with respect to the position of the reception point. A large mouth opening and a large cavity close to the mouth will favor effects of HOM on directivity at relatively low frequency (about 4 kHz). A small mouth opening and a large cavity distant from the mouth exit will induce a reduced influence of the HOM on the directivity. However, in any case, the HOM will influence the transfer function at relatively low frequency (about 4 kHz) by inducing anti-resonances and additional resonances.

Chapter 6

Conclusions and perspectives

6.1 Conclusions

With respect to the objectives outlined in the introduction, the following conclusions are made.

6.1.1 Influence of the appearance of 3D acoustic field in the vocal tract

The work presented in this document allowed one to improve the understanding of the influence of the 3D aspects of the acoustic field on the transfer function of the vocal tract and on the directivity of the radiated sound. It has been observed that the propagation of HOM, which induce 3D variations of the acoustic field, generate anti-resonances and additional resonances. The bandwidth of the resonances has been observed to be affected by the difference in radiation efficiency of the HOM involved. Small changes (about 2%) of the resonance frequencies below the cut-on frequency of the first HOM due to the excitation of evanescent HOM have also been observed.

As expected, the propagation of HOM has been observed to influence consequently the directivity of the radiated sound. Thus, the transfer function has important variations (up to 50 dB) within small frequency intervals (of the order of 100 Hz) and within small angular regions (about 30°). In the case of circular cross-sections, the propagation of the first HOM ψ_{01} generates a two lobes directivity pattern which cannot be predicted by the plane piston model. The lobes of the patterns generated by the HOM can have significant variation of direction (up to 60°) within small frequency intervals (of the order of 100 Hz). It has been observed that the directivity can be locally reduced by the propagation of HOM in comparison with the directivity predicted with the plane mode only.

The comparison of three vocal tract geometrical approximations for the vowels [ɑ], [i] and [u] showed that the effects of HOM vary a lot with the vocal tract geometry. Even though the effects of HOM have been observed in narrow frequency intervals from 3.5 kHz, the expected limit of about 4-5 kHz for the appearance of this phenomenon has been confirmed.

6.1.2 Accurate measurements

The design of an experimental setup, especially dedicated to measure pressure fields and transfer functions inside waveguides, allowed one to perform accurate measurements (with a maximal variability of the order of 1 dB). Directivity measurements with increased frequency and angular resolution (5 Hz and 3°) compared to the previous measurements performed with speakers and singers (third of octave bands and 15°) allowed one to make more accurate observations of the directivity patterns and of their variation with the frequency.

6.1.3 Validation of the 3D acoustic simulation methods

The comparison of the experimental measurements of pressure fields and PPTF with the MM and FEM simulations showed that the acoustic field is successfully predicted by both methods. The frequency of the peaks of the PPTF is predicted with a typical accuracy of 5%. However, the -3 dB bandwidth of the peaks is underestimated by the MM because visco-thermal losses are not taken into account in this implementation. The bandwidths obtained with FEM are closer to the experimental ones but still have significant differences because the losses are considered constant for all the frequencies. Average differences between the pressure fields are of the order of 1dB. On the other hand, even though the resonances and anti-resonances cannot be directly observed with PPTF, their characteristics can be deduced from the observation of the peaks and dips of the PPTF. The projection of the pressure field measured in the plane perpendicular to the propagation axis x_3 showed that there is a good agreement between the modal amplitudes predicted by the MM simulation and the experimental ones. The directivity patterns are successfully predicted by the radiation model implemented, and the effect of a reflection on an object is qualitatively reproduced by a simple model of specular reflection.

6.1.4 Comparison of the MM and the FEM

The FEM can be applied to any geometry, but require long simulation durations (typically of the order of 24h on serial computing system with processor Intel® Core™ i5 2.8 GHz for the geometries studied in this document) and needs to mesh outer space. The MM has a shorter simulation duration (typically of the order of 1 h with processor Intel® Core™ Vpro™ 3.0 GHz for the geometries studied in this document) and does not require any other discretization than the vocal tract, but is limited to unbended straight geometries in the current implementation used.

6.1.5 Effect of geometrical features

The eccentricity appears to be a key factor for the excitation and the propagation of the HOM featuring nodal lines located on a symmetry axis. It has been shown that a very small asymmetry is sufficient to allow these HOM to contribute to the acoustic field, indeed

MM simulations have shown that a shift of 0.5 mm of the centers of the sections of the two tube geometry is sufficient to generate significant effects of the HOM. This is the origin of larger discrepancies between simulation and experiment for the axisymmetric geometries: the replicas used for the experiment being not perfectly axisymmetric, HOM can be slightly excited. At high frequency (above about 5 kHz), the axisymmetric geometries appear as particular theoretical cases difficult to reproduce experimentally which does not correspond to the features of a realistic vocal tract. Thus, the eccentricity is important to take into account for reproducing qualitatively the acoustical properties of the vocal tract when the 3D aspects of the acoustic field need to be taken into account.

The cross-section shape appears as a very important factor which determines the HOM (mode-shapes and cut-on frequencies) potentially present in the geometries. The frequency of the resonances, the coupling between the propagation modes at the change of cross-section and the directivity patterns are determined by the cross-sectional shape.

The size of the mouth aperture influences the directivity and the bandwidth of the resonances. A large mouth aperture is likely to induce larger bandwidths, and directivity effect of the HOM spreads on large frequency intervals. A small mouth aperture is likely to induce small bandwidths and directivity effects limited to small frequency intervals (of the order of 500 Hz) but still present.

The presence of the lips lowers the resonance frequencies and reduces their bandwidth. They appear as a key element in the radiation of the sound by creating low amplitude directions in the vertical plane (x_1,x_2) present at all the frequencies.

6.1.6 Comparison of simplified vocal tract replicas with realistic ones

Building a vocal tract geometry from an area function using elliptical cross-sections and junctions eccentric in both vertical and horizontal planes allows one to get acoustical properties qualitatively close to the ones of a realistic geometry. The use of such kind of geometry would allow one to perform, as an example, sound synthesis based on a physical model with less complexity for the description of the vocal tract geometry and at a lower computational cost.

6.1.7 Comparison of the vowels [ɑ], [i] and [u]

It has been observed with MM simulation that, as expected, the directivity patterns can vary a lot with the considered phoneme, and a physical explanation has been proposed for the difference observed. Indeed, the different directivity properties of the vowels [ɑ], [i] and [u] can be related to the shape of the corresponding geometries. In particular, the effect of the mouth aperture investigated with a two tubes vocal tract approximation is confirmed. A large cavity located close to the mouth induces the appearance of the effect of the HOM at relatively low frequency (about 4 kHz) and if this same cavity is placed closer to the glottis the effect of the HOM appear at higher frequency because the 3D aspects of the acoustic field cannot

be transmitted through the narrower part which connects it to the mouth.

6.2 Perspectives

Based on these conclusions, the following perspectives are proposed.

The effects of the HOM at high frequency (above 5 kHz) need to be taken into account in any physical modeling of vocal tract acoustics and more generally in high frequency studies related to speech and singing. In particular, some effects not present at low frequency such as the presence of anti-resonances and, overall the variation of the transfer function with the position of the reception point need to be considered. Given the significant variations of the HOM within small frequency intervals, one can question how their effect is perceived by a listener. Thus, the perceptual consequences of the HOM need to be further investigated. On the other hand, it would be interesting to pursue this work with geometries corresponding to other phonemes. In particular, the fricatives and the plosives, which have more energy in the high frequencies, and which have been observed to have a pronounced directivity would be very interesting to study. It would be also very interesting to study these effects on real speakers by performing more accurate directivity measurements with a higher frequency and angular resolution and focusing on single phonemes.

Appendix A

Repeatability test 2-10 kHz

In this appendix an experiment complementary to the one detailed in section 3.3.3 is presented. A measurement has been repeated three times with sinus signals at a fixed position on the exit surface of a uniform tube. The frequency has been varied by steps of 10 Hz from 2 kHz to 10 kHz.

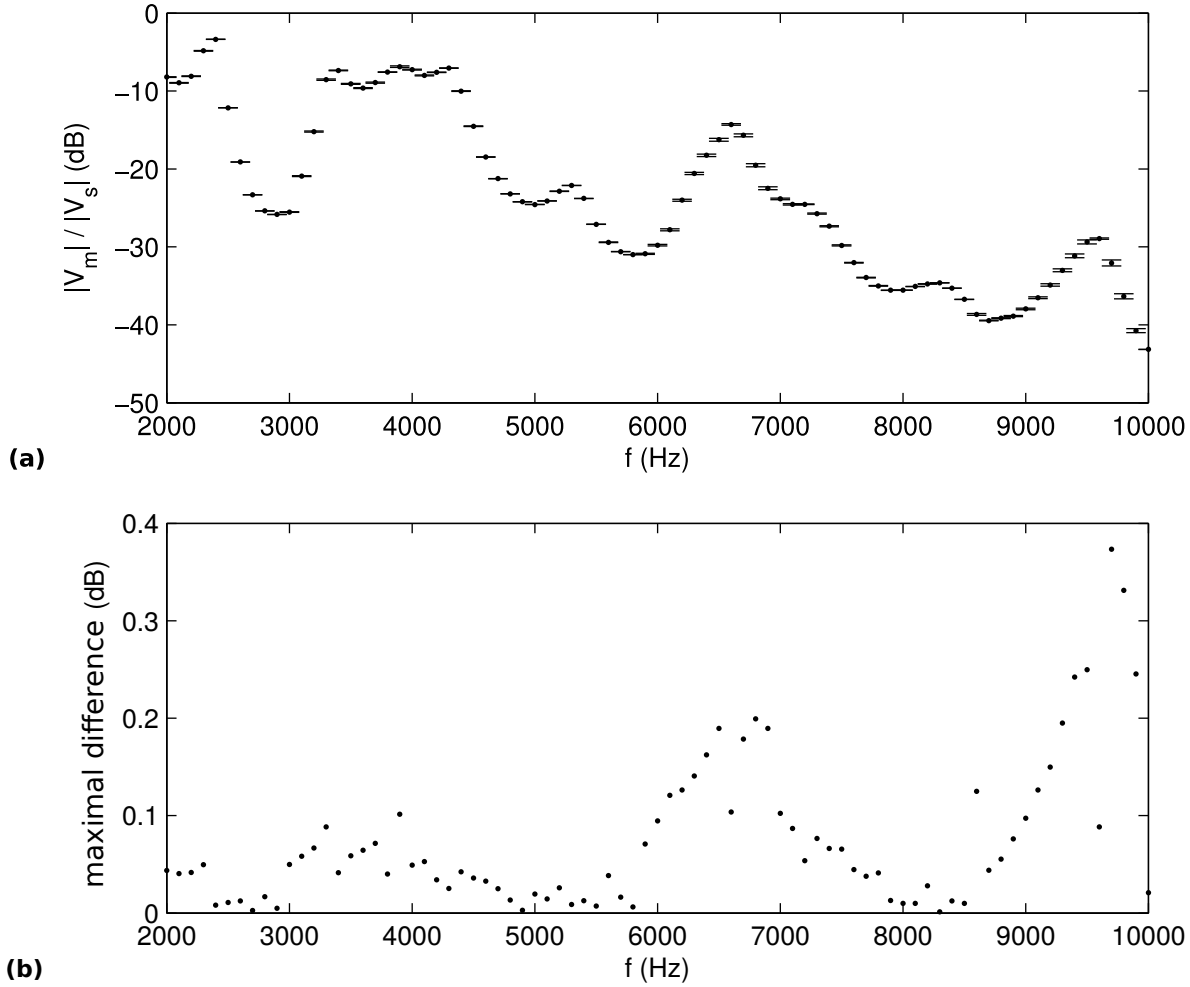


Figure A.1: (a) Amplitude averaged on three measurements performed at the same location (b) maximal amplitude difference between the three measurements

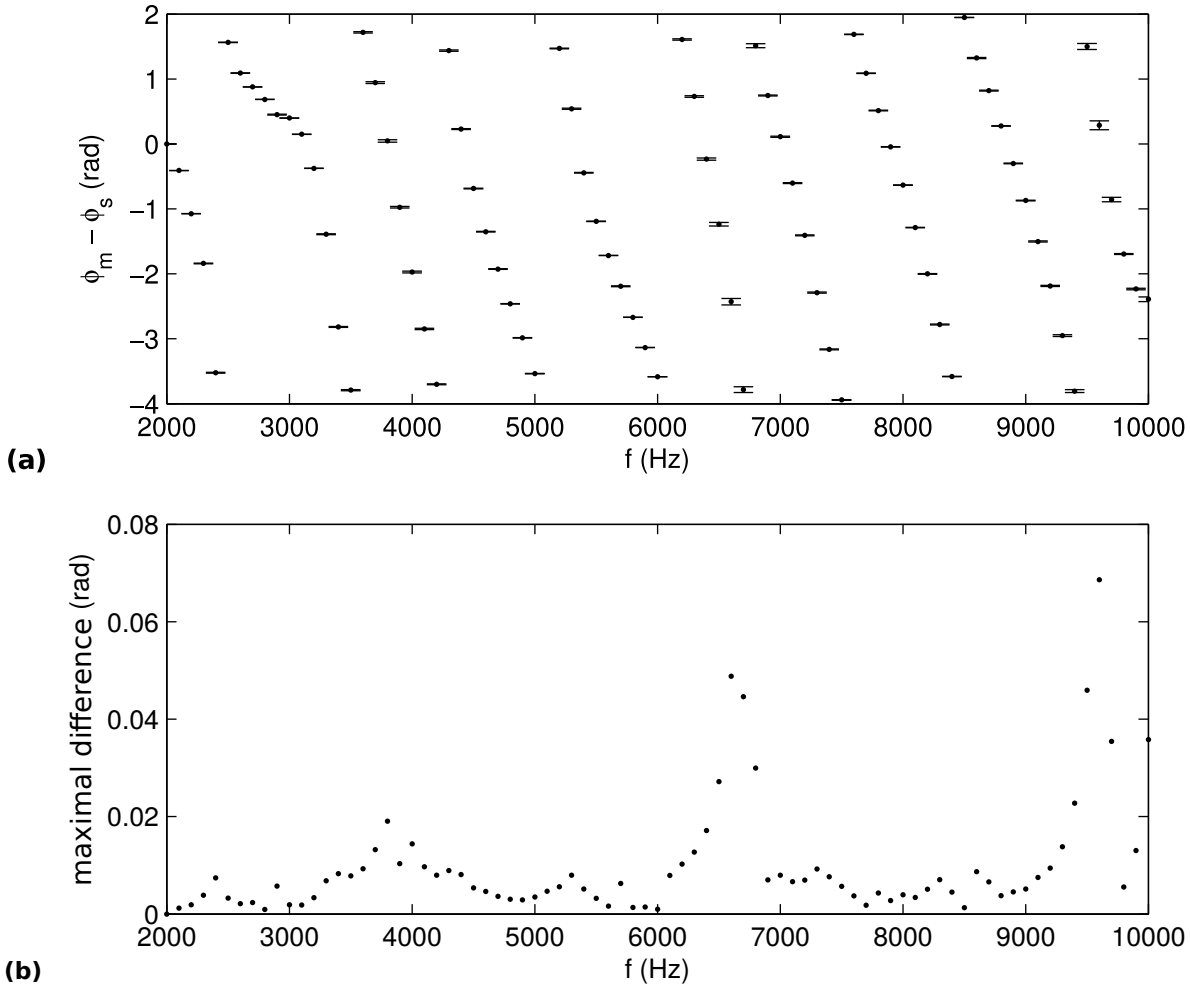


Figure A.2: (a) Phase averaged on three measurements performed at the same location (b) maximal phase difference between the three measurements

The ratio $\frac{|V_m|}{|V_s|}$ is presented as a function of the frequency in Fig. A.1a. It is proportional to the amplitude of the acoustic pressure. As for the experiments presented in section 3.3.3, the variations of amplitude with the frequency can be attributed to the combination of the effect of the acoustic resonances of the tube and the frequency dependence of the sound source amplitude. Error bars show the range of variations of the measured amplitude. The difference between the maximal and the minimal amplitudes for each frequency is presented in Fig. A.1b. One can see that the variations of amplitude over the three experiments remains lower than 0.4 dB. These variations change with the frequency. They can be very low at some frequencies (0.0013 dB at 8.3 kHz), and reach maxima at others (0.038 dB at 9.7 kHz). Globally the variations tend to increase with the frequency. This can be understood as the consequence of the global decrease of the amplitude of the measured acoustic pressure when the frequency is increased. Indeed, the sound source becomes less efficient at high frequency, and the probe of the microphone by damping the high frequencies acts as a low pass filter. As a consequence,

the signal to noise ratio is reduced and more variations can be induced by the measurement noise.

The phase difference $\phi_m - \phi_s$ is presented as a function of the frequency in Fig. A.2a. It corresponds to the phase variations of the acoustic pressure with the frequency. Error bars show the range of variations of the measured phase difference. The difference between the maximal and the minimal phase for each frequency is presented in Fig. A.2b. One can see that the variations of phase between the three experiments remain lower than 0.08 rad. As for the amplitude, it changes with the frequency and can have very low values (down to 0.001 rad at 6 kHz) and reachs maxima at other frequencies (up to 0.07 rad at 9.6 kHz).

In conclusion, the variations of amplitude and phase are respectively lower than 0.4 dB and 0.08 rad for this experiment. This having been carried out within less time (40 minutes) than the experiment with the sinus presented in section 3.3.3 (two hours), the variations due to the change of experimental conditions are less important, and indeed the observed variations are smaller for this experiment (0.9 dB and 0.2 rad for the other experiment). However, there are less repetitions and it is still possible that the effect of the measurement noise on the variations of the measured quantities is less visible. Nevertheless, the smooth variations of the maximal difference can still be interpreted as the effect of the variation of the experimental conditions during the measurement process.

Appendix B

Normalization constants of the propagation modes in circular cross-sections

In order to compute the normalization constant N_{mn} present in the expression of the propagation modes of a circular cross-section (Eq. (2.6)), one applies the normalisation condition of Eq. (2.5):

$$\int_S \frac{J_n(\frac{r\gamma_{mn}}{R})^2}{N_{mn}^2}\cos(n\theta)^2 dS = S \ . \tag{B.1}$$

If $n = 0$, $\cos(n\theta) = 1$ and Eq. (B.1) can be integrated in the following way:

$$\frac{1}{N_{mn}^2}\left[\frac{1}{2}r^2\left(J_n(\frac{r\gamma_{mn}}{R})^2 - J_{n-1}(\frac{r\gamma_{mn}}{R})J_{n+1}(\frac{r\gamma_{mn}}{R})\right)\right]_0^R [\theta]_0^{2\pi} = \pi R^2 \ , \tag{B.2}$$

which results in

$$N_{mn}^2 = \sqrt{J_n(\gamma_{mn})^2 - J_{n-1}(\gamma_{mn})J_{n+1}(\gamma_{mn})} \ . \tag{B.3}$$

If $n > 0$, Eq. (B.1) can be integrated in the following way:

$$\frac{1}{N_{mn}^2}\left[\frac{1}{2}r^2\left(J_n(\frac{r\gamma_{mn}}{R})^2 - J_{n-1}(\frac{r\gamma_{mn}}{R})J_{n+1}(\frac{r\gamma_{mn}}{R})\right)\right]_0^R \left[\frac{\theta}{2} + \frac{\sin(2n\theta)}{4n}\right]_0^{2\pi} = \pi R^2 \ , \tag{B.4}$$

which results in

$$N_{mn}^2 = \sqrt{\frac{1}{2}\left(J_n(\gamma_{mn})^2 - J_{n-1}(\gamma_{mn})J_{n+1}(\gamma_{mn})\right)} \ . \tag{B.5}$$

Note that if a sinus is used in Eq. (2.6), the case $n = 0$ can not be treated because $\psi_{m0} = 0$ and the case $n > 0$ can be integrated and also leads one to the Eq. (B.5).

Appendix C

Detailed computation of the four junction types

One has $(j\boldsymbol{D_3})^{-1} = \boldsymbol{D_2^{-1}D_1}$ and $\boldsymbol{D_1^2 - I = D_2^2}$

C.1 Two consecutives expansions

$$\boldsymbol{Z^{(0)} = (D_1 Z^{(1)} + D_2 Z_c)(D_2 Z_c^{-1} Z^{(1)} + D_1)^{-1}}$$

One introduces the factor $\boldsymbol{D_2^{-1} Z_c}$ everywhere

$$\boldsymbol{Z^{(0)} = (D_2^{-1} Z_c D_2 Z_c^{-1} D_1 Z^{(1)} + D_2 Z_c D_2^{-1} Z_c D_2 Z_c^{-1})(D_2 Z_c^{-1} Z^{(1)} + D_1)^{-1} D_2 Z_c^{-1} D_2^{-1} Z_c}$$

$$\boldsymbol{Z^{(0)} = ((jD_3)^{-1} D_2 Z^{(1)} + D_2^{-1} D_2^2 Z_c)[D_2^{-1} Z_c (D_2 Z_c^{-1} Z^{(1)} + D_1)]^{-1} D_2^{-1} Z_c}$$

$$\boldsymbol{Z^{(0)} = ((jD_3)^{-1} D_2 Z^{(1)} + (D_1^2 - I) D_2^{-1} Z_c)[Z^{(1)} + (jD_3)^{-1} Z_c]^{-1} D_2^{-1} Z_c}$$

$$\boldsymbol{Z^{(0)} = ((jD_3)^{-1} D_2 Z^{(1)} + (jD_3)^{-1} Z_c D_1 - D_2^{-1} Z_c) A}$$

with $\boldsymbol{A = (D_2 Z_c^{-1} Z^{(1)} + D_1)^{-1} = [Z^{(1)} + (jD_3)^{-1} Z_c]^{-1} D_2^{-1} Z_c}$

$$\boldsymbol{A^{-1} = D_2 Z_c^{-1} (Z^{(1)} + (jD_3)^{-1} Z_c)}$$
$$\boldsymbol{A^{-1} = D_2 Z_c^{-1} Z^{(1)} + D_2 (jD_3)^{-1}}$$

$$\boldsymbol{Z^{(0)} = (jD_3)^{-1} Z_c (D_2 Z_c^{-1} Z^{(1)} + D_1 D_2^{-1} D_2) - D_2^{-1} Z_c) A}$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}(\boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + (j\boldsymbol{D_3})^{-1}\boldsymbol{D_2}) - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c})\boldsymbol{A}$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}\boldsymbol{A}^{-1} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c})A$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}A$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}[\boldsymbol{Z}^{(1)} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}$$

C.2 Expansion followed by a contraction

$$\boldsymbol{Z}^{(0)} = (\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})(\boldsymbol{D_2}\boldsymbol{Z_c}^{-1} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)})^{-1}$$

$$\boldsymbol{Z}^{(0)} = (\boldsymbol{D_2}^{-1}\boldsymbol{D_1}\boldsymbol{D_2} + \boldsymbol{D_2}^2\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})(\boldsymbol{D_2}\boldsymbol{Z_c}^{-1} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)})^{-1}\boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}$$

$$\boldsymbol{Z}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{D_2} + (\boldsymbol{D_1}^2 - \boldsymbol{I})\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})[\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}(\boldsymbol{D_2}\boldsymbol{Z_c}^{-1} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)})]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}$$

$$\boldsymbol{Z}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{D_2} + \boldsymbol{D_2}^{-1}\boldsymbol{D_1}^2\boldsymbol{Z_c}\boldsymbol{Y}^{(1)} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})[\boldsymbol{I} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}$$

$$\boldsymbol{Z}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{D_2} + (j\boldsymbol{D_3})^{-1}\boldsymbol{D_1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})\boldsymbol{A}$$

with $\boldsymbol{A} = (\boldsymbol{D_2}\boldsymbol{Z_c}^{-1} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)})^{-1} = [\boldsymbol{I} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}$

$$\boldsymbol{Z}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}[\boldsymbol{Z_c}^{-1}\boldsymbol{D_2} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)}] - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})\boldsymbol{A}$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}\boldsymbol{A}^{-1}\boldsymbol{A} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}\boldsymbol{A}$$

$$\boldsymbol{Z}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}[\boldsymbol{I} + (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}$$

C.3 Two consecutive contractions

$$\boldsymbol{Y}^{(0)} = (\boldsymbol{D_2}\boldsymbol{Z_c}^{-1} + \boldsymbol{D_1}\boldsymbol{Y}^{(1)})(\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})^{-1}$$

$$\boldsymbol{Y}^{(0)} = (\boldsymbol{D_2}^{-1}\boldsymbol{D_2}^2\boldsymbol{Z_c}^{-1} + \boldsymbol{D_2}^{-1}\boldsymbol{D_1}\boldsymbol{D_2}\boldsymbol{Y}^{(1)})(\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})^{-1}\boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

$$\boldsymbol{Y}^{(0)} = (\boldsymbol{D_2}^{-1}(\boldsymbol{D_1}^2 - \boldsymbol{I})\boldsymbol{Z_c}^{-1} + \boldsymbol{D_2}^{-1}\boldsymbol{D_1}\boldsymbol{D_2}\boldsymbol{Y}^{(1)})[\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}(\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

$$\boldsymbol{Y}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{D_1}\boldsymbol{Z_c}^{-1} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1} + (j\boldsymbol{D_3})^{-1}\boldsymbol{D_2}\boldsymbol{Y}^{(1)})[(j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} + \boldsymbol{Y}^{(1)}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

$$\boldsymbol{Y}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1}[\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)}] - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1})\boldsymbol{A}$$

with $\boldsymbol{A} = (\boldsymbol{D_1} + \boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{Y}^{(1)})^{-1} = [(j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} + \boldsymbol{Y}^{(1)}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$

$$\boldsymbol{Y}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{A}^{-1}\boldsymbol{A} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{A}$$

$$\boldsymbol{Y}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}[(j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} + \boldsymbol{Y}^{(1)}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

C.4 Contraction followed by an expansion

$$\boldsymbol{Y}^{(0)} = (\boldsymbol{D_2}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_1})(\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c})^{-1}$$

$$\boldsymbol{Y}^{(0)} = (\boldsymbol{D_2}^{-1}\boldsymbol{D_2}^2\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}^{-1}\boldsymbol{D_1}\boldsymbol{D_2})(\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c})^{-1}\boldsymbol{D_2}\boldsymbol{Z_c}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

$$\boldsymbol{Y}^{(0)} = (\boldsymbol{D_2}^{-1}(\boldsymbol{D_1}^2 - \boldsymbol{I})\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + (j\boldsymbol{D_3})^{-1}\boldsymbol{D_2})[\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}(\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c})]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

$$\boldsymbol{Y}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{D_1}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + (j\boldsymbol{D_3})^{-1}\boldsymbol{D_2})\boldsymbol{A}$$

with $\boldsymbol{A} = (\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c})^{-1} = [\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}(\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c})]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$

$$\boldsymbol{Y}^{(0)} = ((j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1}[\boldsymbol{D_1}\boldsymbol{Z}^{(1)} + \boldsymbol{D_2}\boldsymbol{Z_c}] - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)})\boldsymbol{A}$$

$$\boldsymbol{Y}^{(0)} = (j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1} - \boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)}[(j\boldsymbol{D_3})^{-1}\boldsymbol{Z_c}^{-1}\boldsymbol{Z}^{(1)} + \boldsymbol{I}]^{-1}\boldsymbol{D_2}^{-1}\boldsymbol{Z_c}^{-1}$$

Bibliography

[1] G. Fant, *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. Walter de Gruyter The Hague, 1971, vol. 2, Chap. 1. (cit. on pp. 2, 5, 30, 32, 33, 53, 54, 68).

[2] J. Flanagan, *Speech analysis: Synthesis and perception*. Springer-Verlag Berlin Heidelberg, 1972, Chap. 3 (cit. on p. 2).

[3] D. O'Shaughnessy, *Speech communication: human and machine*, ser. Addison-Wesley series in electrical engineering: digital signal processing. Wiley-IEEE Press Piscataway, 1987, Chap. 3 (cit. on p. 2).

[4] K. Stevens, *Acoustic phonetics*. MIT press Cambridge, 2000, vol. 30, Chap. 3 (cit. on p. 2).

[5] J. Smith, *Waveguide filter tutorial*. Citeseer, 1987 (cit. on p. 2).

[6] H. Dunn and D. Farnsworth, "Exploration of pressure field around the human head during speech", *J. Acoust. Soc Am.*, vol. 10, no. 3, pp. 184–199, 1939 (cit. on p. 2).

[7] T. Halkosaari, M. Vaalgamaa, and M. Karjalainen, "Directivity of artificial and human speech", *J. Audio Eng. Soc.*, vol. 53, no. 7/8, pp. 620–631, 2005 (cit. on p. 2).

[8] D. Cabrera, P. Davis, and A. Connolly, "Long-term horizontal vocal directivity of opera singers: effects of singing projection and acoustic environment", *J Voice*, vol. 25, no. 6, e291–e303, 2011 (cit. on p. 2).

[9] A. Marshall and J. Meyer, "The directivity and auditory impressions of singers", *Acta Acust united Ac*, vol. 58, no. 3, pp. 130–140, 1985 (cit. on p. 2).

[10] B. Monson, E. Hunter, and B. Story, "Horizontal directivity of low-and high-frequency energy in speech and singing", *J. Acoust. Soc Am.*, vol. 132, no. 1, pp. 433–441, 2012 (cit. on pp. 2, 3).

[11] J. Flanagan, "Analog measurements of sound radiation from the mouth", *J. Acoust. Soc Am.*, vol. 32, no. 12, pp. 1613–1620, 1960 (cit. on p. 2).

[12] F. McKendree, "Directivity indices of human talkers in English speech", in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Institute of Noise Control Engineering, vol. 1986, 1986, pp. 911–916 (cit. on p. 2).

[13] J. Peng, T. Wang, and S. Wu, "Investigation on the effects of source directivity of chinese speech intelligibility in real and virtual rooms", *Appl Acoust*, vol. 74, no. 8, pp. 1037–1043, 2013 (cit. on p. 2).

[14] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments", *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999 (cit. on p. 2).

[15] B. Monson, E. Hunter, A. Lotto, and B. Story, "The perceptual significance of high-frequency energy in the human voice", *Frontiers in psychology*, vol. 5, 2014 (cit. on p. 2).

[16] S. El-Masri, X. Pelorson, P. Saguet, and P. Badin, "Development of the transmission line matrix method in acoustics applications to higher modes in the vocal tract and other complex ducts", *Int J Numer Model El*, vol. 11, no. 3, pp. 133–151, 1998 (cit. on p. 3).

[17] T. Vampola, J. Horácek, A. Laukkanen, and J. Švec, "Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement", *Logopedics Phoniatrics Vocology*, no. 0, pp. 1–10, 2013 (cit. on p. 3).

[18] M. Arnela and O. Guasch, "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method", *J. Acoust. Soc Am.*, vol. 133, no. 6, pp. 4197–4209, 2013 (cit. on pp. 3, 34, 51).

[19] K. Motoki, N. Miki, and N. Nagai, "Measurement of sound-pressure distribution in replicas of the oral cavity", *J. Acoust. Soc Am.*, vol. 92, no. 5, pp. 2577–2585, 1992 (cit. on pp. 3, 35).

[20] R. Blandin, M. Arnela, R. Laboissière, X. Pelorson, O. Guasch, A. Van Hirtum, and X. Laval, "Effects of higher order propagation modes in vocal tract like geometries", *J. Acoust. Soc Am.*, vol. 137, no. 2, pp. 832–843, 2015 (cit. on pp. 3, 53).

[21] J. Dang and K. Honda, "Acoustic characteristics of the piriform fossa in models and humans", *J. Acoust. Soc Am.*, vol. 101, no. 1, pp. 456–465, 1997 (cit. on p. 3).

[22] H. Takemoto, P. Mokhtari, and K. T., "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method", *J. Acoust. Soc Am.*, vol. 128, no. 6, pp. 3724–3738, 2010 (cit. on p. 3).

[23] B. Story, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002", *J. Acoust. Soc Am.*, vol. 123, p. 327, 2008 (cit. on pp. 3, 5, 30, 32, 33, 65, 67, 68, 95, 96, 101, 103).

[24] S. Narayanan, A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging", *J. Acoust. Soc Am.*, vol. 98, no. 3, pp. 1325–1347, 1995 (cit. on p. 3).

[25] J. Woo, E. Murano, M. Stone, and J. Prince, "Reconstruction of high-resolution tongue volumes from MRI", *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 12, pp. 3511–3524, 2012 (cit. on p. 3).

[26] D. Aalto, O. Aaltonen, R. Happonen, P. Jääsaari, A. Kivelä, J. Kuortti, J. Luukinen, J. Malinen, T. Murtola, R. Parkkola, *et al.*, "Large scale data acquisition of simultaneous MRI and speech", *Appl Acoust*, vol. 83, pp. 64–75, 2014 (cit. on pp. 3, 4, 34, 76, 77, 99).

[27] A. Roure, "Propagation guidée. etude des discontinuités (Guided propagation. Study of discontinuities)", PhD thesis, Université d'Aix-Marseille, 1976 (cit. on pp. 4, 7).

[28] J. Kergomard, A. Garcia, G. Tagui, and J. Dalmont, "Analysis of higher order mode effects in an expansion chamber using modal theory and equivalent electrical circuits", *J. Sound Vib.*, vol. 129, no. 3, pp. 457–475, 1989 (cit. on pp. 4, 7).

[29] V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation", *J. Acoust. Soc Am.*, vol. 100, no. 4, pp. 2034–2048, 1996 (cit. on pp. 4, 7).

[30] J. Kemp, "Theoretical and experimental study of wave propagation in brass musical instruments", PhD thesis, University of Edinburgh, 2002 (cit. on pp. 4, 7, 15, 16, 19, 20, 22).

[31] S Félix, J. Dalmont, and C. Nederveen, "Effects of bending portions of the air column on the acoustical resonances of a wind instrument", *J. Acoust. Soc Am.*, vol. 131, no. 5, pp. 4164–4172, 2012 (cit. on p. 4).

[32] K. Motoki, P. Badin, X. Pelorson, and H. Matsuzaki, "A modal parametric method for computing acoustic characteristics of three-dimensional vocal tract models", in *Proceedings of 5th Seminar on Speech Production: Models and Data*, 2000, pp. 325–328 (cit. on p. 7).

[33] P. Morse and K. Ingard, *Theoretical acoustics*. Princeton University Press, 1986, Chap. 9. (cit. on pp. 8, 58).

[34] M. Lowson and S. Baskaran, "Propagation of sound in elliptic ducts", *J. Sound Vib.*, vol. 38, no. 2, pp. 185–194, 1975 (cit. on p. 11).

[35] K. Hong and J. Kim, "Natural mode analysis of hollow and annular elliptical cylindrical cavities", *J. Sound Vib.*, vol. 183, no. 2, pp. 327–351, 1995 (cit. on p. 11).

[36] N. McLachlan, "Theory and application of Mathieu functions", 1947 (cit. on p. 11).

[37] P. Morse, H. Feshbach, *et al.*, *Methods of theoretical physics (Parts 1 and 2)*, 2. McGraw-Hill New York, 1953 (cit. on p. 11).

[38] M. Abramowitz and I. Stegun, *Handbook of mathematical functions*, 5. Dover New York, 1972, vol. 1, chap. 20. (cit. on p. 12).

[39] I. Singer and E. Turkel, "High-order finite difference methods for the Helmholtz equation", *Computer Methods in Applied Mechanics and Engineering*, vol. 163, no. 1, pp. 343–358, 1998 (cit. on p. 12).

[40] M. McClure, "Images of a vibrating Koch drum", *Computers & Graphics*, vol. 32, no. 6, pp. 711–715, 2008 (cit. on p. 12).

[41] W. Zorumski, "Generalized radiation impedances and reflection coefficients of circular and annular ducts", *J. Acoust. Soc Am.*, vol. 54, p. 1667, 1973 (cit. on pp. 15–18).

[42] J. Kemp, D. Campbell, and N. Amir, "Multimodal radiation impedance of a rectangular duct terminated in an infinite baffle", *Acta Acust united Ac*, vol. 87, no. 1, pp. 11–15, 2001 (cit. on p. 15).

[43] V. Fromkin, "Lip positions in American English vowels", *Language and speech*, vol. 7, no. 4, pp. 215–225, 1964 (cit. on p. 34).

[44] S. Dabbaghchian, M. Arnela, and O. Engwall, "Simplification of vocal tract shapes with different levels of detail.", in *Proc. of 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, Scotland, UK, 2015 (cit. on p. 34).

[45] N. Ruty, X. Pelorson, A. Van Hirtum, I. Lopez-Arteaga, and A. Hirschberg, "An in vitro setup to test the relevance and the accuracy of low-order vocal folds models", *J. Acoust. Soc Am.*, vol. 121, no. 1, pp. 479–490, 2007 (cit. on p. 35).

[46] A. Van Hirtum and Y. Fujiso, "Insulation room for aero-acoustic experiments at moderate Reynolds and low Mach numbers", *Appl Acoust*, vol. 73, no. 1, pp. 72–77, 2012 (cit. on p. 37).

[47] J. Lord Rayleigh, *Theory of sound*. Dover two volumes 1945, 1896, vol. 2, 2nd ed., pp. 333-342 (cit. on p. 38).

[48] W. Nyborg, "Acoustic streaming due to attenuated plane waves", *J. Acoust. Soc Am.*, vol. 25, no. 1, pp. 68–75, 1953 (cit. on p. 38).

[49] P. Westervelt, "The theory of steady rotational flow generated by a sound field", *J. Acoust. Soc Am.*, vol. 25, no. 1, pp. 60–67, 1953 (cit. on p. 38).

[50] F. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978 (cit. on p. 38).

[51] O. Guasch and F. Magrans, "The global transfer direct transfer method applied to a finite simply supported elastic beam", *J. Sound Vib.*, vol. 276, no. 1, pp. 335–359, 2004 (cit. on p. 48).

[52] P. Švancara and J. Horáček, "Numerical modelling of effect of tonsillectomy on production of Czech vowels", *Acta Acust united Ac*, vol. 92, no. 5, pp. 681–688, 2006 (cit. on p. 51).

[53] M. Arnela, O. Guasch, and F. Alías, "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations", *J. Acoust. Soc Am.*, vol. 134, no. 4, pp. 2946–2954, 2013 (cit. on p. 51).

[54] M. Arnela, R. Blandin, S. Dabbaghchian, O. Guasch, F. Alías, X. Pelorson, A. Van Hirtum, and O. Engwall, "Influence of lips on the production of vowels based on finite element simulations and experiments", *J. Acoust. Soc Am.*, vol. 139, no. 5, pp. 2852–2859, 2016 (cit. on pp. 53, 76).

[55] A. Pierce *et al.*, *Acoustics: an introduction to its physical principles and applications*. McGraw-Hill New York, 1981, vol. 20 (cit. on pp. 82, 87).

[56] R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, "Influence of higher order acoustical propagation modes on variable section waveguide directivity: application to vowel [a]", *Acta Acust united Ac*, vol. 102, no. 5, 2016 (cit. on p. 83).

Résumé — L'acoustique du conduit vocal est souvent décrite avec de simples modèles ondes planes à une dimension. Cependant, cela n'est pas satisfaisant quand, à haute fréquence (à partir d'environ 5 kHz), des variations tridimensionnelles du champ acoustique sont présentes. La théorie acoustique multimodale a été implémentée pour prendre en compte les aspects tridimensionnels de l'acoustique du conduit vocal. Un système expérimental a été conçu pour mesurer avec précision des fonctions de transfert, des champs de pression et des diagrammes de directivité sur des maquettes de conduits vocaux. Les données expérimentales obtenues ont été comparées avec les simulations réalisées avec la théorie implémentée et avec la méthode des éléments finis. Le champ acoustique tridimensionnel et les diagrammes de directivité ont été prédis avec succès par les deux méthodes de simulation. Il a été observé que la propagation de mode acoustique d'ordre supérieur induit des variations tridimensionnelles du champ acoustique, génère des antirésonances et des résonances additionnelles, et affecte la directivité du son rayonné de façon significative. L'excentricité de la forme du conduit vocal apparaît comme critique pour l'excitation et la propagation des modes d'ordre supérieur. Il est conclu qu'à haute fréquence (au-delà de 5 kHz), la fonction de transfert du conduit vocal peut avoir des variations significatives dans de petits intervalles de fréquences (de l'ordre de 100 Hz) et dans des régions angulaires restreintes (de l'ordre de 30°). Ces variations nécessitent d'être prises en compte dans les études de la parole qui se focalisent sur les hautes fréquences.

Mots clés : parole, conduit vocal, acoustique, champ acoustique tridimensionnel, méthode multimodale, directivité

Abstract — The vocal tract acoustics is often described with a simple one dimensional plane wave approach. However, this is not satisfying when at high frequency (from about 5 kHz) three dimensional variations of the acoustic field are present. The multimodal acoustic theory has been implemented in order to account for the three dimensional aspects of the vocal tract acoustics. An experimental setup has been designed to measure accurately transfer functions, pressure field maps and directivity patterns of vocal tract replicas. The experimental data obtained have been compared with simulations performed with the implemented theory and with a finite element method. The three dimensional acoustic fields and the directivity patterns were successfully predicted by both simulation methods. It has been observed that the propagation of higher order acoustical modes, induces three dimensional variations of the acoustic field, generates anti-resonances and additional resonances, and significantly affects the directivity of the radiated sound. The eccentricity of the vocal tract shape appears as critical for the excitation and the propagation of the higher order acoustical modes. It is concluded that at high frequency (above 5 kHz), the transfer function of the vocal tract can have significant variations within short frequency intervals (of the order of 100 Hz) and within small angular regions (of the order of 30°) which need to be taken into account in the studies of speech which focus on high frequencies.

Keywords: speech, vocal tract, acoustic, three-dimensional acoustic field, multimodal method, directivity

Gipsa-lab, 11 rue des mathématiques, Grenoble Campus BP46, F-38402
SAINT MARTIN D'HERES CEDEX