Contribution to the three-dimensional analysis of the vocal tract acoustics.

X. Pelorson(1), K. Motoki(2), R. Laboissière(1).

(1) : Institut de la Communication Parlée,
CNRS, INPG, Université Stendhal,
46 avenue F. Viallet
F-38031, Grenoble Cedex 01.
(2) : Faculty of Engineering,
Hokkai-Gakuen University
S-26, W-11, Sapporo, Japan.

Abstract

The acoustics of the vocal tract at high frequencies, including higher modes of propagation, is discussed using three dimensional solutions of the wave equation. Different geometries of increasing complexity are considered. It is shown, in particular, that the effect of the position of the source and of the vocal tract geometry appears to play a crucial role at high frequencies. The application of these findings with respect to vowel and consonant synthesis is discussed.

Introduction.

During the last decades most efforts, in the field of speech modeling, have been focussing on the description of the sound sources while the acoustic propagation within the vocal tract still relies on a one dimensional plane wave assumption (e.g. Fant, 1960). If not clearly established, the limit of validity of such an assumption is already well-known to lie in the range 4-5 kHz which is a severe limitation in particular for plosive and fricative modeling. The use of one dimensional acoustics could have been justified by the lack of information about the vocal tract real geometry. However, nowadays, thanks to new medical imaging techniques such as MRI, accurate three-dimensional representations of the vocal tract geometry are available.

In this paper the 3-D theory of wave propagation inside a duct is briefly discussed. The application of such a theory to the case of a cascade tube representation of the vocal tract, using a mode matching technique, is then presented. The effects of three-dimensional wave propagation are then illustrated using simple synthetic configurations.

I. The one tube approximation of the vocal tract.

We consider in this section, the simplest representation of the vocal tract, illustrated in figure 1, as a uniform tube with rigid walls, closed at one end (the "glottis") and opened in an infinite baffle at the other end (the "lips"). The shape of the vocal tract is considered as rectangular with a cross-section of 2.6 x 2.6 cm and a total length of 16.8 cm. For the sake of simplicity, we only consider that the excitation of the vocal tract is provided by a single point source, Q, located at (x_0, y_0, z_0). Due to the linearity of the acoustical problem, the results presented here can easily be extended to the case of multiple sources or of spatially distributed sources.

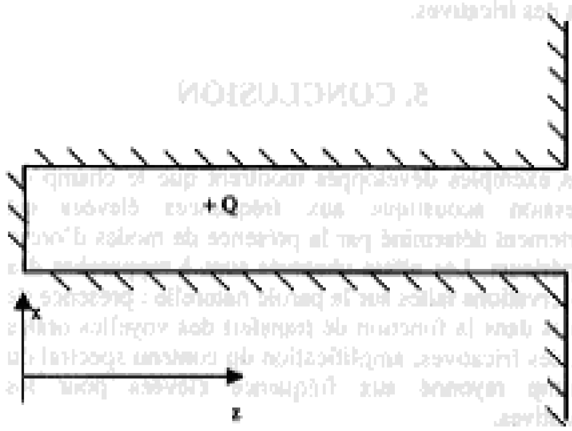

Figure 1: Simple one tube representation of the vocal tract.

Using these assumptions, the general solution of the acoustic wave propagation in the frequency domain is:

$$P(f,x,y,z) = \sum_{m,n} \Lambda_{mn} Q \frac{\Psi_{mn}(x,y)\Psi^*_{mn}(x_0,y_0)}{k_{mn}^2 - k^2} \times \left(A_{mn} e^{-jk_{mn}z} + B_{mn} e^{jk_{mn}z}\right) \quad (1)$$

where f is the frequency imposed by the source, A_{mn} and B_{mn} are two constants determined by the boundary conditions at both ends of the wave guide, Λ_{mn} is a constant, and the eigen function Ψ_{mn} is defined by:

$$\Psi_{mn}(x,y) = \cos\left(\frac{m\pi}{L_x} x\right) \cos\left(\frac{n\pi}{L_y} y\right) \quad (2)$$

The wave constant k_{mn} is determined by the dispersion equation:

$$k_{mn}^2 = k^2 - \left(\frac{m\pi}{L_x}\right)^2 - \left(\frac{n\pi}{L_y}\right)^2 \quad (3)$$

Equation (1) together with equation (3) shows that a given mode (m,n) will be propagating only if the excitation frequency, $f = kc/2\pi$, is higher than the cut on frequency:

$$f_{mn} = \frac{c}{2\pi}\sqrt{\left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2},$$

where c is the speed of sound.

In particular, below the first cut-on frequency f_{01} or f_{10}, all k_{mn}s, m,n ≠ 0, are purely imaginary, the corresponding modes (m,n) are thus decaying and at a finite distance from the source, the three-dimensional solution (1) converges to a one dimensional one (plane wave).

Although equation (1) requires an infinite summation over both m and n, it is not necessary, in practice, to account for a large number of higher modes. As a typical example, we present, in figure 2, the pressure distribution calculated inside the one tube approximation using two higher modes. As a reference, this result is compared to the numerical simulation using the TLM method (El Masri et al., 1998) which includes all higher modes up to the frequency associated to the spatial resolution (of order of 50 kHz).

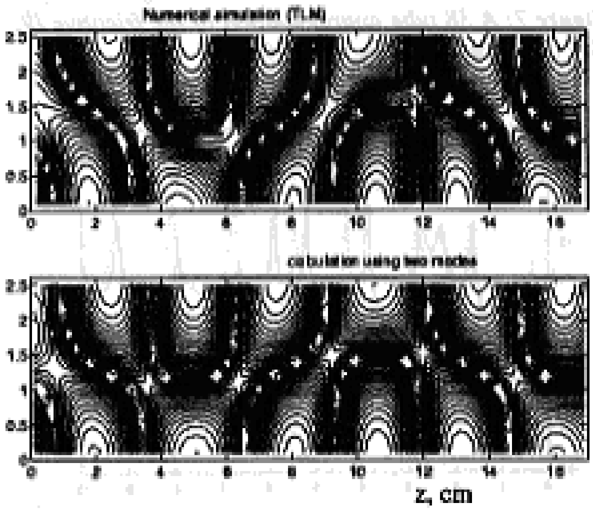

Figure 2 : Pressure contours (in 1 dB step) at 9 kHz obtained by numerical simulation (top) and calculation (bottom).

Equation (1), through the term $\Psi^*_{mn}(x_0, y_0)$, clearly shows that the efficiency of the higher modes is a function of the position of the sound source within the vocal tract. A sound source located at the center of a section ($x_0 = L_x/2$ or $y_0 = L_y/2$), such as the glottis, in first approximation, will not generate any odd-order higher modes because in such a case $\Psi^*_{mn}(x_0, y_0) = 0$. On the opposite, a sound source located near a wall will generate odd-order higher modes with a maximum amplitude ($|\Psi^*_{mn}(x_0, y_0)| = 1$).

II. Effects of a change in geometry, the two-tube approximation.

We now consider another configuration depicted in figure 3. In the first case, the vocal tract is assumed to be symmetrical along the z-axis while in the second one the vocal tract is assumed totally asymmetrical. It is clear that none of these situations are realistic, the real vocal tract shape is expected to be somewhat in between the extreme cases presented in Figure 3.

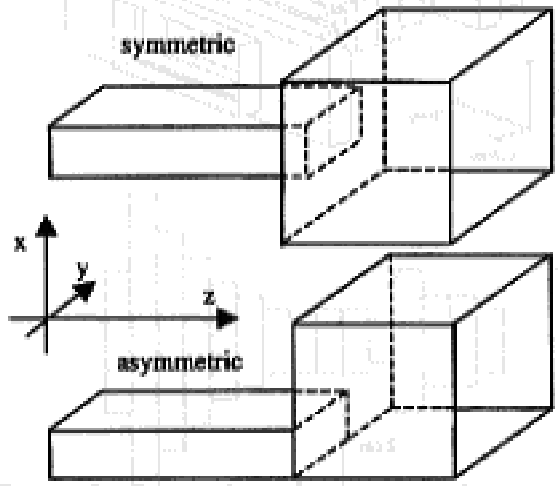

Figure 3: Symmetric versus asymmetric two-tube approximation.

In both cases an analytical solution is, in general, impossible. The problem can however be solved using a numerical resolution based on a matching mode technique (Kergomard, 1991). In figure 4 are presented two typical examples of pressure distributions in the z-y plane, for both symmetrical and asymmetrical 2-tube configurations at 6 kHz. As one would expect, while in the symmetrical case the first higher mode (0,1) is not present, in the asymmetrical case, the influence of this mode can clearly be observed.

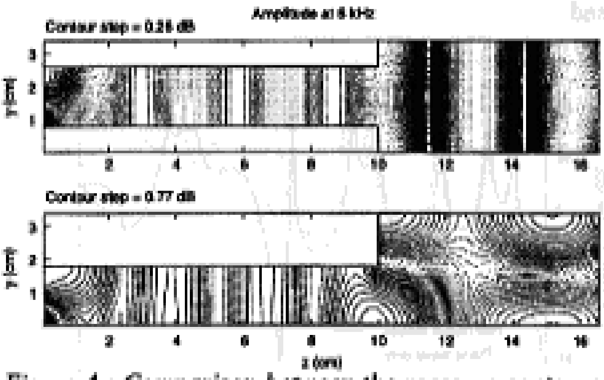

Figure 4 : Comparison between the pressure contours in the y-z plane at 6 kHz obtained with a centric and an eccentric 2-tube approximation.

III. N-tubes approximations.

As two last examples, we present calculations performed for more realistic vocal tract geometries derived from MRI data. In both cases the source is located at (0,0,0).

The first case corresponds to the vowel /a/ modeled as a 15-tube geometry as shown in figure 5, in this case the geometry is imposed as symmetrical along the z-axis.

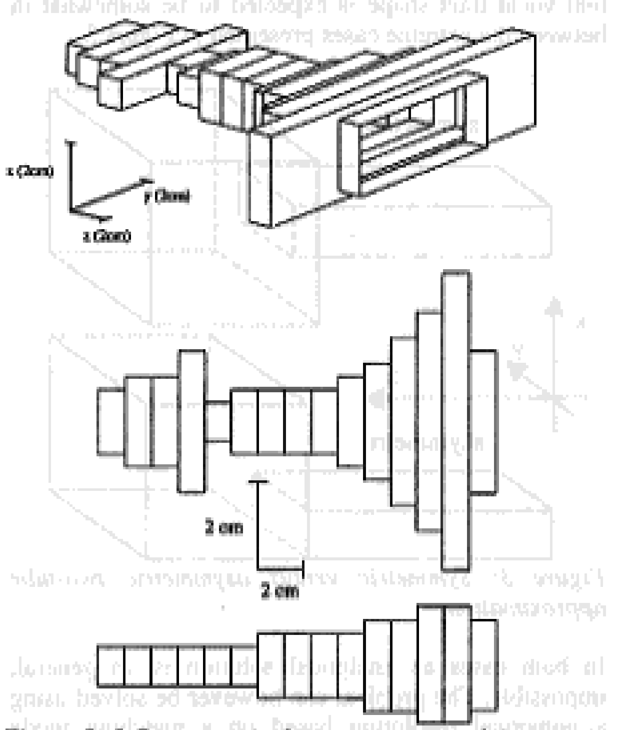

Figure 5: 3-D geometrical approximation of the vowel /a/. Top : 3-D view, bottom : side views.

As amore synthetic representation, we present, in figure 6, the calculated transfer function defined by :

$$H \propto \left|\sqrt{\frac{W_{rad}}{Q}}\right|$$

where W_{rad} is the total power radiated from the open end.

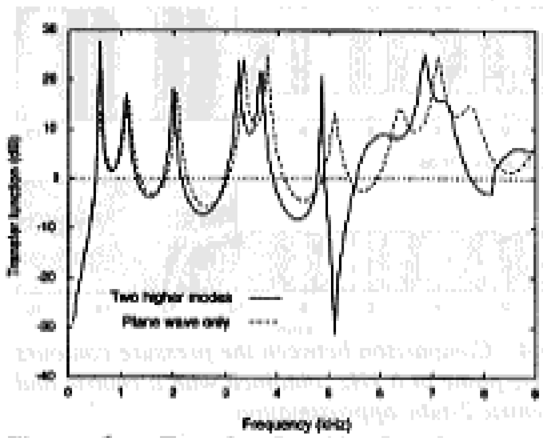

Figure 6 : Transfer function for the vowel /a/. Comparison between plane wave solution and solution including two higher modes.

The last example concerns the fricative consonant /ʃ/ modeled using 38 tubes as illustrated in figure 7. Note that in this last case, an asymmetry is introduced to conform with MRI data. The corresponding transfer function is presented in figure 8.

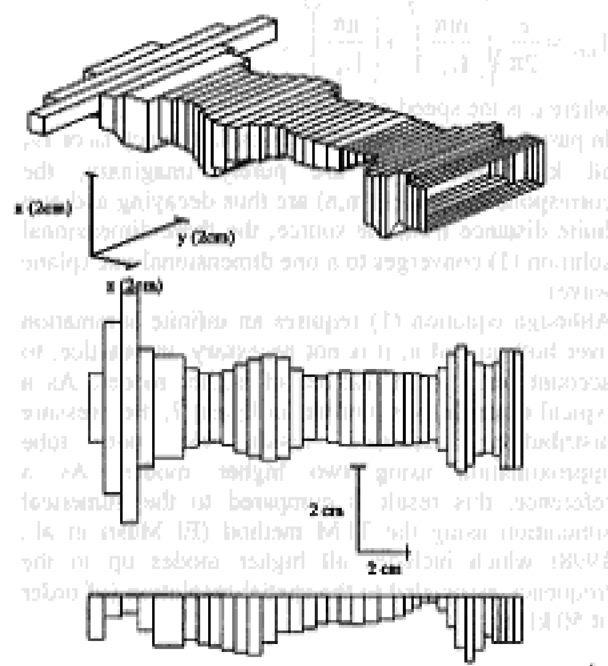

Figure 7: A 38 tube approximation of the fricative /ʃ/. Top : 3-D view, bottom : side views.

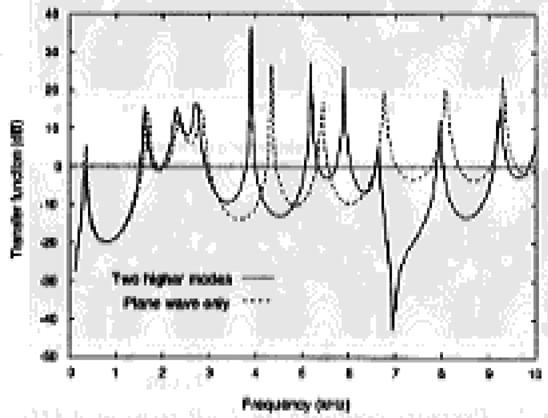

Figure 8 : Transfer function for the fricativel /ʃ/. Comparison between plane wave solution and solution including two higher modes.

As shown in figures 6 and 8, and by comparison with the one-dimensional results (plane wave solution), the presence of higher modes have two effects :

- a reduction of the resonance frequencies below the cut-on frequency of the first higher mode due to the presence of vanishing modes
- a "zero" occurring at the cut-on frequency of the first higher mode and, further, a severe perturbation of the resonance pattern.

Conclusion

The examples presented above clearly show that higher modes of propagation can have a significant effect on vocal tract acoustics. When compared with real speech data, these effects can be related with some well-known phenomenon such as the presence of zeros in the transfer function of oral vowels or the zeros and the strong enhancement of energy at high frequency for fricatives.

An accurate description of higher modes effects is however difficult because, as illustrated in this paper, the efficiency of these modes depends not only on the position of the source but also on the precise geometry of the vocal tract. More precisely, the example presented in section II shows that the degree of asymmetry of the vocal tract plays a crucial role. The real vocal tract is, of course, not as asymmetrical as assumed in section II but certainly not perfectly symmetrical. Further research, coupled with 3-D geometrical modeling of the vocal tract, is thus needed before one can draw a clear conclusion.

References

Elmasri S., Pelorson X., Saguet P., Badin P. (1998). "The use of the Transmisson Line Matrix in acoustics and in Speech", International Journal of Numerical Modelling, 11, 133-151.

Fant G. (1970) "Acoustic theory of speech production", 2nd Ed., Mouton and co, Den Haag.

Kergomard, J. (1991) "Calculation of discontinuities in waveguides using mode-matching method : an alternative to the scattering matrix approach", J. Acoustique, 4, 111-138.

Acknowledgements:

This work has been partially supported by the French CNRS federation of INPG laboratories ELESA