A Quasi-One-Dimensional Model of Aerodynamic and Acoustic Flow in the Time-Varying Vocal Tract: Source and Excitation Mechanisms

Gordon Ramsay

Laboratoire de Phonologie
Université Libre de Bruxelles
Brussels, Belgium

Abstract

In this paper, the conservation laws of classical fluid mechanics are used to derive a quasi-one-dimensional model of fluid flow in an elastic tube of time-varying cross-sectional area, representing the human vocal tract. The global flow equations are then decomposed into aerodynamic and acoustic components, representing respiratory flow and sound propagation during speech. The nature of the coupling between the two systems is investigated, and a new interpretation of the traditional source-filter model of speech production is proposed.

1. Introduction

Speech is produced by establishing a respiratory flow in the vocal tract, which causes the vocal folds to vibrate, and may create turbulent regions within the fluid. Both of these effects can excite the resonances of the time-varying oral, tracheal, and nasal cavities to produce the acoustic flow perturbations that are perceived as sound, although the mechanisms responsible for the generation of acoustic sources are not yet well understood.

Respiratory flow is generally neglected in traditional acoustic models of the vocal tract. However, if this is done, it is difficult to locate, within the physical equations governing the flow, the source of external energy that must drive the acoustic system if sound is to be produced. The traditional solution, on which the source-filter model is based, is to cut off the vocal tract at the level of the glottis, and to specify the "source" in terms of a predefined volume velocity waveform applied at the entry to the oral tract, which is assumed to act as a simple linear filter. It is not clear why the vocal tract should terminate at the glottis, rather than the lungs, and the form of the supposed glottal source waveform has never been satisfactorily explained.

Classical aeroacoustics on the other hand indicates that sound production should be linked to coupling between the acoustic field and non-acoustic entropy and vorticity gradients in the respiratory flow [1]. A number of recent attempts have been made to derive proper aeroacoustic models of sound production, where the generation of acoustic sources by aerodynamic phenomena is taken into account, based mainly on consideration of three-dimensional vortex motion [2] [3] [4] [5].

The present paper demonstrates that a source-filter model, including aerodynamic effects, can also be derived directly from the quasi-one-dimensional conservation laws, using modal analysis. A number of alternative non-vortical mechanisms are proposed by which sources may arise within the time-varying vocal tract. Notably, by decomposing the global flow field into an incompressible aerodynamic flow and an isentropic acoustic flow (cf. [6]), it is shown that source terms may arise naturally from aerodynamic flow gradients, and from dynamic coupling between different eigenmodes due to rapid area variations.

2. Conservation laws

Suppose that the vocal tract can be modelled as an elastic tube of length L and time-varying cross-sectional area $A : \Omega \to \mathbb{R}_+$, defined on a bounded rectangle $\Omega = \{(x, t) : x \in [0, L], t \in [0, T]\}$ in $\mathbb{R}^2$, where x represents the distance along the tract mid-line from the lungs to the lips, and t represents time.

Assume that the physical state of the air within the tube can be described by functions $\rho, p, s, u, e, \theta : \Omega \to \mathbb{R}$, representing respectively the time-varying air density, pressure, entropy, particle velocity, specific internal energy and temperature along the tube length, averaged over the tube cross-section. If body forces, rotational motion, heat transfer and viscous effects are neglected, the principles of conservation of mass, momentum, and energy, applied to the fluid within the tube, can be used to derive the following system of partial differential equations, which must be obeyed everywhere on the interior of Ω;

$$\frac{\partial}{\partial t}\rho A + \frac{\partial}{\partial x}\rho A u = 0, \quad (1)$$

$$\frac{\partial}{\partial t}\rho A u + \frac{\partial}{\partial x}\rho A u^2 + A\frac{\partial p}{\partial x} = 0, \quad (2)$$

$$\frac{\partial}{\partial t}(\rho e + \tfrac{1}{2}\rho u^2)A + \frac{\partial}{\partial x}(\rho e + \tfrac{1}{2}\rho u^2 + p)A u = 0. \quad (3)$$

Together with an appropriate set of boundary conditions, the quasi-one-dimensional conservation laws (1)-(3) describe the evolution of the global flow field within the time-varying tube. It is important to realize that the global flow field does not, in general, represent only the acoustic perturbations that define the sound field, but may well include other components of fluid motion that arise from non-acoustic phenomena. Sound propagation is intrinsically associated with compressible isentropic (adiabatic and reversible) perturbations of the fluid flow, but incompressible or non-isentropic fluctuations may also be supported by the movement of the fluid, and are not excluded as solutions of the equations of motion. The respiratory flow established during speech, for example, is not acoustic in nature, since it does not propagate or radiate into the far field as sound. In order to explain the physical mechanisms that are specifically responsible for the generation of sound in the vocal tract, a means must therefore be found of separating the acoustic and non-acoustic, or aerodynamic, components of the flow field and of characterizing their interaction.

Before this can be done, it is necessary to specify precisely how the aerodynamic and acoustic components of the global flow field are to be defined. In the fluid mechanics literature, the quasi-one-dimensional conservation laws are often analysed by linearizing the global flow equations about a steady-state solution that may vary in space but not in time. It is sometimes possible to suppose that the aerodynamic field can be defined

by the steady-state solution, and that the acoustic field corresponds to a simple linear perturbation superimposed thereon. However, this is not guaranteed to capture the isentropic character of sound propagation, and cannot be adopted here, since movement of the vocal tract walls will evidently induce spatio-temporal fluctuations in the flow field that may be both aerodynamic and acoustic in nature. An alternative approach will be proposed instead, based on decomposing the pressure field into compressible isentropic and incompressible non-isentropic time-varying perturbations about an equilibrium state. This extends an earlier analysis undertaken by Jospa [6] (cf. [1]).

3. Aeroacoustic decomposition

To introduce an aeroacoustic decomposition of the global flow field, assume that the fluid within the vocal tract is initially at rest in some state $(\rho_\star, p_\star, s_\star, u_\star)$, and that the physical state of the fluid at any subsequent time can be represented as the sum of an aerodynamic perturbation $(\Delta\rho_0, \Delta p_0, \Delta s_0, \Delta u_0)$ and an acoustic perturbation $(\Delta\rho_1, \Delta p_1, \Delta s_1, \Delta u_1)$, superimposed on the initial state; denoting by $(\epsilon_\rho, \epsilon_p, \epsilon_s, \epsilon_u)$ any residual approximation error, which is assumed to be second-order,

$$\rho = \rho_\star + \Delta\rho_0 + \Delta\rho_1 + \epsilon_\rho, \tag{4}$$

$$p = p_\star + \Delta p_0 + \Delta p_1 + \epsilon_p, \tag{5}$$

$$s = s_\star + \Delta s_0 + \Delta s_1 + \epsilon_s, \tag{6}$$

$$u = u_\star + \Delta u_0 + \Delta u_1 + \epsilon_u. \tag{7}$$

The equations governing the global flow field are highly non-linear, and it is not possible to assume that the two flow components are independent, or that any principle of linear superposition can be employed. However, if the aerodynamic and acoustic fields are assumed to be weakly coupled, and generated by different physical mechanisms, it is often possible to construct a unique decomposition of this kind by imposing physically-meaningful constraints on the evolution of the two components.

Since sound propagation in air is known to be isentropic, it is useful to begin by examining the transport of entropy within the fluid. Expanding the energy equation (3), and eliminating terms using equations (1) and (2), it is not difficult to show that

$$\rho A \left(\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} \right) + p \frac{\partial}{\partial x} A u = 0. \tag{8}$$

The second law of thermodynamics provides a differential relationship linking the entropy s to ρ, p, e, and θ,

$$\theta \, ds = de - \frac{p}{\rho^2} d\rho, \tag{9}$$

and if the air in the vocal tract is assumed to behave as an ideal gas, the following constitutive law also holds, where R is the gas constant,

$$p = \rho R \theta. \tag{10}$$

Substituting equations (9) and (10) into equation (8) yields the entropy conservation law for quasi-one-dimensional flow;

$$\frac{\partial s}{\partial t} + u \frac{\partial s}{\partial x} = R \frac{\partial}{\partial t} \ln A. \tag{11}$$

This is an interesting result that is often overlooked. When the area function is static, the term on the right disappears, entropy fluctuations are convected with the flow, and the global

flow field itself is isentropic. However, in the more general case, temporal variations in the tube cross-section can be seen to induce local entropy gradients that are proportional to the temporal derivative of the logarithm of the area function. Given that the acoustic field is, by definition, isentropic, these entropy variations must be associated with the non-acoustic component of the flow field. This immediately suggests that the aerodynamic flow component (defining Δs_0) should be chosen to obey equation (11) directly, whereas the acoustic flow component (defining Δs_1) must necessarily satisfy the homogeneous version of equation (11) in which the source term is set to zero.

To proceed further, recall that any single thermodynamic variable can always be expressed as a function of any two others. The pressure p can thus be written as a function $p(s, \rho)$ of the entropy s and the density ρ. Assuming that this function is continuously-differentiable, any first-order fluctuations in pressure Δp can be expressed in terms of equivalent first-order fluctuations in density, $\Delta\rho$, and entropy, Δs, by expanding the pressure in a Taylor series about the equilibrium state:

$$\Delta p = \left. \frac{\partial p}{\partial s} \right|_{\rho_\star} \Delta s + \left. \frac{\partial p}{\partial \rho} \right|_{s_\star} \Delta\rho + O(\Delta s^2, \Delta\rho^2). \tag{12}$$

The first term represents the partial pressure fluctuation due to entropy changes with the density held constant, whereas the second term represents the partial pressure fluctuation due to density changes with the entropy held constant. Comparing (12) with (5), since the acoustic component is isentropic by definition, this implies that Δp_1 must be an *isentropic* pressure change, corresponding to the second term in (12), with $\Delta s_1 = 0$, whereas Δp_0 must be an *incompressible* pressure change, corresponding to the first term in (12), with $\Delta\rho_0 = 0$:

$$\Delta p_0 := \left. \frac{\partial p}{\partial s} \right|_{\rho_\star} \Delta s_0, \tag{13}$$

$$\Delta p_1 := c_\star^2 \, \Delta\rho_1, \tag{14}$$

where $c_\star$ can be shown to be the local sound speed,

$$c_\star := \sqrt{ \left. \frac{\partial p}{\partial \rho} \right|_{s_\star} }. \tag{15}$$

The aerodynamic and acoustic density, pressure, and entropy components of the fluid flow have now been defined; it only remains to assume that the motion of the fluid can be represented in the same manner, by supposing that the global particle velocity u can meaningfully be decomposed into an aerodynamic particle velocity Δu_0 associated with $(\Delta\rho_0, \Delta p_0, \Delta s_0)$ on which an acoustic particle velocity Δu_1 associated with $(\Delta\rho_1, \Delta p_1, \Delta s_1)$ is superimposed. Since incompressible respiratory flow and isentropic acoustic flow in speech propagate through different mechanisms, and evolve on very different time scales, this is not an unreasonable assumption to make.

This completes the desired aeroacoustic decomposition. To derive the equations that govern the evolution of the aerodynamic and acoustic state variables, it will be convenient to describe the flow field in terms of dimensionless groups $A_0, \alpha_0, \rho_0, P_0, U_0, S_0, \rho_1, P_1, U_1, S_1, V_1 : \omega \to \mathbb{R}$, defined on a domain $\omega = \{(\xi, \tau) : \xi \in [0, 1], \tau \in [0, c_\star T/L]\}$, where $\tau := c_\star t/L, \xi := x/L, A_0 := A/L^2$, and

$$\begin{aligned}
\rho_0 &:= \Delta\rho_0/\rho_\star, & \rho_1 &:= \Delta\rho_1/\rho_\star, \\
P_0 &:= \Delta p_0/\rho_\star c_\star^2, & P_1 &:= \Delta p_1/\rho_\star c_\star^2, \\
S_0 &:= \Delta s_0/R, & S_1 &:= \Delta s_1/R, \\
U_0 &:= \Delta u_0/c_\star, & U_1 &:= \Delta u_1/c_\star, \\
\alpha_0 &:= \ln A_0, & V_1 &:= A_0 U_1.
\end{aligned}$$

Remark that U_0 and U_1 are the convective and acoustic Mach numbers, and $\rho_0 = 0$, $s_1 = 0$, $p_1 = \rho_1$ by construction. It will also be convenient to introduce the differential operators

$$\frac{D}{D\tau} := \frac{\partial}{\partial\tau} + U_0 \frac{\partial}{\partial\xi}, \qquad (16)$$

$$\frac{D}{D\xi} := \frac{\partial}{\partial\xi}. \qquad (17)$$

When U_0 is constant, $D/D\tau$ and $D/D\xi$ define a frame of reference that is convected with the aerodynamic flow.

4. Aerodynamic model

By substituting (4)–(7) into equations (1)–(3), and neglecting terms of first order or higher, the following system of equations is obtained, describing the aerodynamic state of the fluid:

$$\frac{\partial A_0}{\partial\tau} + \frac{\partial}{\partial\xi} A_0 U_0 = 0, \qquad (18)$$

$$\frac{\partial U_0}{\partial\tau} + \frac{\partial}{\partial\xi}(P_0 + \tfrac{1}{2}U_0^2) = 0, \qquad (19)$$

$$\frac{\partial S_0}{\partial\tau} + U_0 \frac{\partial S_0}{\partial\xi} = \frac{\partial\alpha_0}{\partial\tau}. \qquad (20)$$

Equation (18) describes conservation of volume, whereas equation (19) is the Bernoulli equation for unsteady incompressible flow. Expressed in terms of the symbolic operators (16) and (17), these can be given a particularly simple form:

$$\frac{DU_0}{D\xi} = Q_u, \qquad (21)$$

$$\frac{DU_0}{D\tau} + \frac{DP_0}{D\xi} = 0, \qquad (22)$$

$$\frac{DS_0}{D\tau} = Q_s, \qquad (23)$$

where Q_u is the mass source represented by

$$Q_u := -\frac{D\alpha_0}{D\tau}, \qquad (24)$$

and Q_s is the entropy source given by

$$Q_s := \frac{\partial\alpha_0}{\partial\tau}. \qquad (25)$$

As boundary conditions, assume that the fluid is initially at rest,

$$\tau = 0 \,:\, P_0(\xi, 0) = U_0(\xi, 0) = S_0(\xi, 0) = 0; \qquad (26)$$

and that a time-varying pressure P_L is applied at the lungs, with atmospheric pressure maintained at the lips,

$$\xi = 0 \,:\, P_0(0, \tau) = P_L(\tau), \qquad (27)$$

$$\xi = 1 \,:\, P_0(1, \tau) = 0. \qquad (28)$$

The aerodynamic equations are parabolic, and can be solved directly once the time-varying area function is specified.

5. Acoustic model

By substituting (4)–(7) into equations (1)–(3), eliminating terms using the aerodynamic equations (18)–(20), and neglecting terms of second order or higher, the following system of equations is obtained, describing the acoustic state of the fluid:

$$\frac{DP_1}{D\tau} + \frac{1}{A_0}\frac{DV_1}{D\xi} + P_1\left(\frac{DU_0}{D\xi} + \frac{D\alpha_0}{D\tau}\right) = 0, \qquad (29)$$

$$\frac{DV_1}{D\tau} + A_0\frac{DP_1}{D\xi} + V_1\left(\frac{DU_0}{D\xi} - \frac{D\alpha_0}{D\tau}\right) = Q_p, \qquad (30)$$

$$\frac{DS_1}{D\tau} = 0, \qquad (31)$$

where Q_p is the momentum source given by

$$Q_p := -\rho_1 A_0 \frac{DU_0}{D\tau}. \qquad (32)$$

As boundary conditions, assume that the fluid is initially at rest,

$$\tau = 0 \,:\, P_1(\xi, 0) = U_1(\xi, 0) = S_1(\xi, 0) = 0; \qquad (33)$$

and that the vocal tract is closed at the entrance to the lungs, and open at the lips, neglecting radiation effects,

$$\xi = 0 \,:\, U_1(0, \tau) = 0, \qquad (34)$$

$$\xi = 1 \,:\, P_1(1, \tau) = 0. \qquad (35)$$

The acoustic equations are hyperbolic, with coefficients that depend on the area function and on the aerodynamic flow gradients. As is often the case in aeroacoustics, the interpretation of the source term Q_p is problematic, since it depends on ρ_1. Following the analogy introduced by Lighthill, the justification for defining it to be an acoustic source rests on the observation that the left-hand side of equations (29)-(31) evidently defines an acoustic wave operator, the symmetry of which would be broken if Q_p were included. If the source term is omitted, the model exhibits a useful duality property: the equations of conservation of mass and momentum can each be obtained from the other by substituting P_1 for V_1, and $1/A_0$ for A_0.

6. Modal analysis

Within the limits of the aeroacoustic assumptions that have been introduced, the acoustic equations essentially describe a linear time-varying system, with coefficients that are modulated by the area function as well as by the aerodynamic flow derivatives. The space of all possible solutions for any particular time-varying area function can therefore be described locally as a linear superposition of characteristic modal vibrations, each of which corresponds to an acoustic resonance, or formant, of the vocal tract. The problem of characterizing the behaviour of the acoustic system in terms of the time-varying area function and aerodynamic flow field reduces to that of determining the spectrum of eigenvalues and eigenfunctions of the corresponding time-varying linear operator (cf. [7] [6]). Individual formant oscillations can subsequently be characterized as the projection of the acoustic state trajectory onto the invariant subspaces associated with each of the time-varying eigenmodes.

To make this precise, define acoustic state and source functions $z := (P_1, V_1, S_1)$, and $q := (0, Q_p, 0)$, which are assumed to dwell in the Hilbert space of square-integrable functions on ω, equipped with an inner-product $\langle\cdot, \cdot\rangle$, and observe that equations (29)-(31) may be re-written in operator form as

$$\dot{z}_t = \mathcal{L}_t z_t + q_t, \qquad (36)$$

where $\mathcal{L}_t$ is the appropriate linear time-varying operator, which will be assumed to be normal, and which depends on A_0 and the aerodynamic flow derivatives. Let $\{\lambda_t^i\}$ be the spectrum of eigenvalues of $\mathcal{L}_t$ with eigenfunctions $\{\phi_t^i\}$, and suppose that $\mathcal{L}_t$ possesses an adjoint $\mathcal{L}_t^\star$ with eigenvalues $\{\overline{\lambda_t^i}\}$ and eigenfunctions $\{\psi_t^i\}$. Under the assumption that the eigenvalues of $\mathcal{L}_t$ are distinct, the eigenfunctions of $\mathcal{L}_t$ and $\mathcal{L}_t^\star$ can be chosen to form a complete biorthogonal basis for the state space, with:

$$\langle \psi_t^i, \phi_t^j \rangle = \delta_{ij}. \qquad (37)$$

Denoting by e_t^i the projection of z_t onto the invariant subspace of $\mathcal{L}_t$ associated with λ_t^i, the following decomposition holds:

$$e_t^i = \langle \psi_t^i, z_t \rangle, \qquad (38)$$

$$z_t = \sum_i e_t^i \phi_t^i, \qquad (39)$$

and by substituting (38) and (39) into equation (36),

$$\dot{e}_t^i = \lambda_t^i e_t^i + \langle \psi_t^i, q_t \rangle - \sum_{j \neq i} \langle \psi_t^i, \dot{\phi}_t^j \rangle e_t^j. \qquad (40)$$

The original infinite-dimensional system of partial differential equations describing the spatio-temporal evolution of the entire acoustic field can thus be reduced to a countable collection of ordinary differential equations describing the temporal evolution of the modal projections. The eigenvalues λ_t^i are the time-varying complex poles of the acoustic system, and define the instantaneous formant frequencies and bandwidths of the modelled vocal tract. The eigenfunctions ϕ_t^i represent the instantaneous spatial distribution of pressure, flow rate, and entropy associated with each formant, whereas the adjoint eigenfunctions ψ_t^i determine the proportion of energy entering each eigenmode from an arbitrary spatial distribution of mass, momentum, and entropy sources. The eigenvalues occur in complex conjugate pairs, and each eigenmode essentially behaves as a simple harmonic oscillator, driven by a source term that depends on the temporal derivative of the aerodynamic flow, and a coupling term that transfers energy from other eigenmodes according to the temporal eigenfunction derivatives, which evidently disappear when the area function is static.

7. Source-filter interpretation

Equations (21)-(23) and (29)-(31) define a complete aerodynamic and acoustic model of quasi-one-dimensional flow in the time-varying vocal tract. The aerodynamic equations clearly demonstrate that the effect of temporal variations in the area function is to establish time-varying flow and entropy gradients within the vocal tract. The acoustic equations reduce to Webster's horn equation when there is no aerodynamic flow, and can be converted to an equivalent bank of formant oscillators with time-varying centre frequencies and bandwidths by projecting the acoustic field onto a dynamic eigenmode basis.

In the absence of aerodynamic flow, there are no acoustic sources. The presence of spatial aerodynamic flow gradients (which can only arise from temporal area function variations) introduces a linear acoustic flow resistance (momentum loss) that increases during the formation of a constriction. The presence of temporal aerodynamic flow gradients (which may arise from temporal area function variations or from the boundary conditions) introduces non-linear coupling between the aerodynamic and acoustic flow fields that behaves like a momentum source, or acoustic dipole, distributed along the vocal tract.

Temporal area function variations also induce fluctuations in the eigenvalues and eigenfunctions of the acoustic system, which induce an additional apparent source that is related to the temporal eigenfunction derivatives. Numerical simulations carried out previously using an artificial point source to replace the aerodynamic source distribution suggest that formant excitation is dominated by the eigenmode coupling terms, and not by the aerodynamic source term itself [8]. The aerodynamic sources appear to provide an amorphous supply of low-frequency energy that is then shaped by the movement of the glottis, via the eigenmode coupling mechanism, into the highly-structured wideband excitation that is observed during voiced speech..

It is interesting to compare this with the traditional source-filter model, where the vocal tract is modelled as a linear time-invariant system driven by an acoustic monopole defined by the aerodynamic flow derivative at the glottis. The derivation presented here suggests that the vocal tract can indeed be represented as a bank of formant oscillators, but that the traditional interpretation of the source may be incorrect. The aerodynamic source term is an acoustic dipole, of magnitude $\rho_1 A_0 DU_0/D_\tau$, and not an acoustic monopole of magnitude $D(\rho_0 A_0 U_0)/D_\tau$, and this does not provide the dominant formant excitation.

8. Conclusions

Aerodynamic and acoustic models describing respiratory flow and sound propagation in a time-varying vocal tract have been derived from the global conservation laws governing quasi-one-dimensional fluid flow. By projecting the solution space of the acoustic system onto a time-varying biorthogonal eigenmode basis, it is possible to decompose any state trajectory into its component formant oscillations. Using this technique, it has been shown that each formant appears to be driven by an excitation composed of an external source, arising from temporal aerodynamic flow gradients, and an internal source arising from temporal eigenfunction modulations that occur when the area function changes rapidly in time, or equivalently when there are large spatial aerodynamic flow gradients, e.g. due to glottal motion. This interpretation is consistent with the traditional source-filter model of speech production, but the nature and origin of the proposed source mechanism are entirely different.

9. References

[1] Goldstein, M. E. (1976). *Aeroacoustics*, McGraw-Hill, New York.

[2] McGowan, R. S. (1988). An aeroacoustic approach to phonation, *J. Acoust. Soc. Am.* **83**(2): 696–704.

[3] Hirschberg, A. (1992). Some fluid dynamic aspects of speech, *Bulletin de la Communication Parlée* **2**: 7–30.

[4] Barney, A., Shadle, C. H. & Davies, P. O. A. L. (1999). Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory, *J. Acous. Soc. Am.* **105**(1): 444–455.

[5] Krane, M. H., Sinder, D. J. & Flanagan, J. L. (2000). Aeroacoustic modeling of speech sound production, *Proc. 5th Seminar on Speech Production*, Kloster Seeon, pp. 177–180.

[6] Jospa, P. (1977). *Théorie acoustique du conduit vocal de forme variable dans le temps*, Ph.D. thesis, U.L.B., Brussels.

[7] Meyer-Eppler, W. & Ungeheuer, G. (1957). Die Vokalartikulation als Eigenwertproblem, *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* **10**(3): 245–257.

[8] Ramsay, G. (1998). An analysis of modal coupling effects during the glottal cycle : formant synthesizers from time-domain finite-difference simulations, *Proc. ICSLP '98*, Sidney, pp. 539–542.