Acoustic tubes with maximal and minimal resonance frequencies

B. De Boer
Spuistraat 210, 1012VT Amsterdam, Netherlands
b.g.deboer@uva.nl
This paper presents a theoretical derivation of acoustic tract shapes (modeled after the vocal tract and similar acoustic horns) that minimize and maximize resonance frequencies. The derivation is based on a symmetry of Webster’s horn equation and on Ehrenfest’s adiabatic invariance hypothesis. It is shown that for minimizing resonance frequencies, abrupt transitions in the area of the tract are necessary, while for maximizing resonance frequencies, gradual transitions are needed. The shape for the tract with the minimal resonance frequency is a tube with constant maximal area near the closed end, followed by a tube of equal length and constant minimal area near the open end. The shape for the tract with maximal resonance frequency consists of a tube with constant minimal area near the closed end, connected to an equally long tube of maximal area through an exponential horn that has the resonance frequency as its cutoff frequency. Tracts for higher resonances can be constructed by concatenating these basic tracts.

1 Introduction

This paper presents the derivation of the analytically exact area functions for acoustic tracts with maximal and minimal resonance frequencies. The only assumption is that Webster’s horn equation [1] is a valid approximation. Although the present paper is a purely theoretical endeavor, it aims to increase our understanding of the maximal range of signals that could be produced by biological vocal tracts. Distinctiveness of signals in biology and linguistics is usually measured by comparing the peaks (or formants) in the spectrum. These peaks are the resonances of the vocal tract with which the signals were generated.

Attempts have been made to systematically explore the space of possible formant patterns of the human vocal tract [2] or to determine vocal tracts with certain acoustic properties by successive approximation [e.g. 3, 4]. However, all previous approaches have made assumptions about the allowable shapes of the vocal tract, so that not all possible shapes could be taken into account. Most approaches have looked at concatenated cylindrical tubes [e.g. 5, 6]. Approaches based on a superposition of a small number of sines and cosines [7, 8], or based on a set of perturbation functions, such as the ones used in [9, section B V.6] have also been made. In this paper, no assumptions about the area function under investigation are made.

The core method used in this paper is adiabatic invariance. In the words of Ehrenfest, adiabatic invariance means that: “If a system is exposed to adiabatic influences, the ‘admissible’ motions are transformed into ‘admissible’ ones” [10]. Translated to acoustics, admissible motions are the resonance frequencies, and adiabatic means changes that are slow compared to the resonance frequencies involved. Adiabatic invariance has first been used in the study of the vocal tract by Schroeder [8] who, using Ehrenfest’s theorem, demonstrated that this makes it possible to determine what happens to resonance frequencies of an acoustic tract when it is (slowly) deformed. Here it will be used to determine the (analytically) exact acoustic tract shapes for producing minimal and maximal resonance frequencies.

The present paper is closely related to Carré [3]. Both investigate optimal acoustic tracts and both use adiabatic invariance [8]. However, in Carré’s paper the main interest is in discovering the vocal tract configuration that maximizes the range of speech sounds that can be made, whereas in this paper, individual tubes with maximal or minimal resonances are determined. Also, Carré’s paper uses numerical simulation while the present paper analytically derives a mathematical result.

2 Optimal tracts

When determining acoustic tracts with minimal and maximal resonance frequencies, it is useful to take a symmetry of Webster’s horn equation [1, 11] into account. For the analysis presented here, it is convenient to split this equation into one for pressure, and one for volume velocity. It then becomes [12, equation 3.2]:

\[
\frac{\partial u}{\partial x} = -\frac{A}{\rho c^2} \frac{\partial p}{\partial t} ; \quad \frac{\partial p}{\partial x} = -\rho \frac{\partial u}{\partial t} \tag{1}
\]

where \(u(x,t)\) is volume velocity, \(p(x,t)\) is the differential pressure, \(A(x)\) is the cross-sectional area of the tract at point \(x\) along its length, \(\rho\) is the density of air and \(c\) is the speed of sound. It can be verified by substitution that if \(u_1\) and \(p_1\) are solutions for a tract with area function \(A_1\), then \(u_2\) and \(p_2\) are a solution for a tract with area function \(A_2\) if:

\[
u_2 = C_a p_1, \quad p_2 = \frac{1}{C_a^2} u_1, \quad A_2 = C_b C_a \frac{\rho^2 c^2}{A_1} \tag{2}
\]

where \(C_a\) and \(C_b\) are arbitrary constants.

Now consider a tract that is closed at one end (\(u\) is zero, \(p\) is maximal) and open at the other (\(u\) is maximal, \(p\) is zero) with an area function that fulfills the following relation:

\[
A(L-x) = \frac{A_{\text{min}} A_{\text{max}}}{A(x)} \tag{3}
\]

where \(L\) is the length of the tract and \(A_{\text{min}}\) and \(A_{\text{max}}\) are the minimal and maximal cross-sectional area. Such a tract will be called symmetric with respect to (2), because the pressure in one direction is related to the volume velocity in the other direction as follows:

\[
p(L-x,t) = \frac{\rho c}{\sqrt{A_{\text{min}} A_{\text{max}}}} u(x,t) \tag{4}
\]

Now, for any acoustic tract that has \(u(0,t) = 0\) and \(p(L,t) = 0\), there must be a point in the tract where, for a given resonance frequency:

\[
p(x,t) = \frac{\rho c}{\sqrt{A_{\text{max}} A_{\text{min}}}} u(x,t) \tag{5}
\]

because \(u\) and \(p\) change gradually. If \(x \neq L/2\), it follows from (4) that it is possible to construct a symmetric tract with the same resonance frequency that is either longer (shorter) than the original tract by cutting the original tract in two at point \(x\), taking the longest (shortest) part and replacing the other part with a part that is equally long to the remaining part, such that the total area function becomes symmetric with respect to (3). This process is illustrated in Fig. 1. As a vocal tract with a minimal
Adiabatic invariance is defined by Schroeder’s Eq. 2 [8]:
\[ \frac{\delta f_n}{f_n} = \frac{\delta E_n}{E_n} \]  
(6)
Where \( f_n \) is the frequency of the \( n \)th resonance, and \( E_n \) is its energy. A small change in relative energy causes an equal small change in relative frequency. This relation holds true for many physical systems, and its validity for acoustical tracts was independently derived by Fant [13].

When deforming a tube in which a standing wave exists, energy is added to or subtracted from the wave. This is because (again, following Schroeder) a standing wave exerts radiation pressure. Deforming the tube requires an amount of work given by Schroeder’s equation 3:
\[ \delta E_n = -\int_0^L P_a(x) \delta A(x) \, dx \]  
(7)
where \( L \) is the length of the tube, \( P_a(x) \) is the radiation pressure of the \( n \)th resonance at position \( x \), and \( \delta A(x) \) is the (small) change in cross-sectional area at position \( x \). As only the first resonance will be studied in the mathematical derivation, the subscript \( n \) will be omitted from now on.

Finally, an expression for the acoustic radiation pressure is needed. Here, Lee and Wang’s equation 27 [14] is used:
\[ \langle P \rangle = \frac{\langle p^2 \rangle}{2 \rho c^2} - \frac{\rho \langle v^2 \rangle}{2} \]  
(8)
where \( \langle \ldots \rangle \) indicates the average over time, \( P \) is the radiation pressure, \( \rho \) is the pressure of the wave, and \( v \) is the particle velocity of the wave, \( \rho \) is the density of air at rest and \( c \) is the speed of sound in air at rest. This expression is equivalent to Story’s [4] perturbation function.

### 2.1 Minimization of the first resonance

At the closed end of the tract, pressure is maximal and volume velocity is zero. At the open end, the reverse obtains. From Eq. (8) it follows that radiation pressure is therefore maximal at the closed end and minimal at the open end, illustrated in Fig. 2. From Eqs. (6) and (7) it then follows that, starting from any area function, maximization of area near the closed end and minimization near the open end will result in minimization of energy and therefore in minimization of resonance frequency.

It can therefore be assumed that the optimal tract of length \( L \) consists of a cylindrical tube of length \( l \leq L/2 \) and area \( \mathcal{A}_\text{max} \) at the closed end. Given the symmetry expressed in (3) at the open end the tract must consist of a cylindrical tube of identical length and area \( \mathcal{A}_\text{min} \). Although this reasoning is certainly correct very close to the ends of the tube, it is nevertheless possible that a different situation obtains inside the tube, and that in fact maximizing cross-sectional area will not decrease frequency.

It must therefore be determined what happens to the energy when a thin slice at the end of the tube with maximal area is replaced by a tube with smaller cross-sectional area. By solving the wave equation in the system consisting of a wide tube connected to a narrower tube, the radiation pressure at the start of the narrower tube can be calculated to be:
\[ P_a(x) = P_o \cos^2 kx - \left( \frac{A_{\text{max}}}{a} \right)^2 \sin^2 kx \]  
(9)
where \( P_o(x) \) is the radiation pressure for area \( a \) of the narrower tube, \( P_o \) is the radiation pressure at the closed end, \( x \) is the distance from the closed end, \( \mathcal{A}_{\text{max}} \) is the area of the first tube and \( k \) is the wave number (\( 2\pi/c \)), with \( f \) the frequency.

The amount of work necessary to reduce a small section of length \( \delta \alpha \), from \( \mathcal{A}_{\text{max}} \) to \( a \) is given by integrating over the change in area:
\[ \delta E(x,a) = \delta l \int_{\mathcal{A}_{\text{max}}}^a P_a(x) \, d\alpha \]  
(10)
which, when (9) is used as the radiation pressure, solves to:
\[ \delta E(x,a) = \delta l \cdot P_o \left( \mathcal{A}_{\text{max}} - a \right) \cos^2 kx + \left( \mathcal{A}_{\text{max}} - \frac{\mathcal{A}_{\text{max}}^2}{a} \right) \sin^2 kx \]  
(11)
As \( \delta \) and \( P_o \) are positive by definition, the sign of the change in energy only depends on the terms between brackets. It can be determined that it is equal to zero for the following condition:
\[ \tan^{-1} \sqrt{\frac{a}{\mathcal{A}_{\text{max}}}} = kx \]  
(12)
For larger values of \( kx \), the change in energy is negative, otherwise it is positive. The minimal value of \( kx \) is reached when \( a \) is equal to \( \mathcal{A}_{\text{min}} \). For given \( k \) and \( \mathcal{A}_{\text{min}} \) there is therefore a position \( x \), such that for every point nearer to the closed end, the total energy, and therefore the resonance frequency cannot be reduced by reducing the tract’s cross sectional area.
The question now becomes: how far away from the ends is this point \( x \)? It can be shown that it is exactly halfway the tract. For a tract with area \( A_{\text{max}} \) at the closed end and area \( A_{\text{min}} \) at the open end, the following condition therefore obtains:

\[
\tan^{-1} \sqrt{\frac{A_{\text{min}}}{A_{\text{max}}}} = k_2 \frac{L}{2} \quad (13)
\]

(where \( k_2 \) is the wave number of the two-cylinder tract). This can be proven by induction. The starting point is a cylindrical tube of length \( L \) with constant area \( A_2 \). The lowest resonance of this tube is a quarter wave, and its wave number \( k_2 \) is therefore \( \pi/(2L) \). The value of (12) for \( A_{\text{min}} = A_{\text{max}} = A_2 \) is equal to \( \pi A_2 \), and therefore \( x = L/2 \). The condition holds for the cylindrical tract.

Now suppose there is a tract consisting of two cylinders, the first with area \( A_1 \), and the second with area \( A_2 \). Suppose also that the wave number of the first resonance is \( k_2 \). By solving the acoustic wave equations it can be found that the total energy (the energy density integrated over the volume) in this tract is given by:

\[
E = \frac{2C^2}{\rho c} A_1 L \quad (14)
\]

where \( C \) is a constant that determines the amplitude of pressure at the closed end. Now a small change in area \( \delta A \) of the first cylinder will result in a change in energy of the total tract of:

\[
\delta E = -\int_{\delta A} \mathcal{P}(x) \, dx = -\delta A \frac{2C^2}{\rho c} \sin k_2 L \quad (15)
\]

where the second term was determined by solving the wave equation in the first cylinder. The relative change in energy is therefore:

\[
\frac{\delta E}{E} = -\frac{\delta A \sin k_2 L}{A_1 \frac{2k_2 L}{k_2}} \quad (16)
\]

Which according to the adiabatic hypothesis equals the relative change in wave number \( \delta k_2/k_2 \).

Now, because we assume that \( k_2 \) must have the value given by (13) a small change \( \delta A \) in \( A_1 \), would need to be offset by the following change in \( k_2 \):

\[
\delta k_2 = \frac{2}{L} \tan^{-1} \sqrt{\frac{A_1}{A_1 + A_2}} = -\delta A \frac{\sqrt{A_1/A_2}}{L(A_1 + A_2)} \quad (17)
\]

The relative change can then be written as:

\[
\frac{\delta k_2}{k_2} = -\frac{\delta A A_1}{A_1} \frac{A_2 + A_1}{A_1} \frac{A_1}{kL} \quad (18)
\]

It turns out that, using the value of \( k_2 \) given by (13):

\[
\sqrt{\frac{A_1 A_2}{A_1 + A_2}} = \frac{\sin k_2 L}{2} \quad (19)
\]

so that Eq. (18) is equal to Eq. (16) Therefore the change in \( k_2 \) caused by the increase in energy is exactly equal to the change in \( k_2 \) needed to maintain condition (13). This condition therefore holds for all tubes consisting of two cylinders of equal length.

Now suppose that the tract with the lowest resonance frequency has a different area function than the one found above. This area function necessarily has an area \( A_{\text{small}} < A_{\text{max}} \) somewhere in the first half of the tract. It also has a lower resonance frequency, and therefore a lower wave number \( k' < k_2 \). Given relations (12) and (13) this means that \( x > L/2 \). But this means that the wave number could be lowered by increasing the section with area \( A_{\text{small}} \) to \( A_{\text{max}} \). But this is in contradiction with the assumption that this tract had the lowest possible wave number and resonance frequency. There is therefore no tract with lower energy, and therefore lower resonance frequency, than the one with a cylinder of maximal area as the first half and a cylinder of minimal area as the second half. Its resonance frequency can be determined from Eq. (13).

### 2.2 Maximization of the first resonance

When maximizing resonance frequency, energy must be maximized. Given the boundary conditions, it appears useful to decrease the area of the half of the tract that is near the closed end, and increase the area of the half that is farthest from the closed end. Although the energy does increase in this way, the situation is different from minimization. For minimization, the wave number decreased, and the radiation pressure therefore stayed uniformly positive in the first half of the tract, and uniformly negative in the second half. In the case of maximization, however, the wave number increases, and this causes the sign of the radiation pressure to change at some point in each half of the tract. This is illustrated in Fig. 3. An extra increase in energy can therefore be achieved if area is increased in this part of the first half of the tract, and decreased in the corresponding part of the second half of the tract. This leads to a gradual transition between the first part and the second part.

The first and last parts of the maximizing tract will still have constant area. By substituting \( A_{\text{min}} \) for \( A_{\text{max}} \) in Eq. (11) and inspecting what happens to the change in energy when area is increased (in the part near the closed end) it is found that a similar condition to (12) obtains. However, now energy can only decrease if one increases cross-sectional area nearer to the closed end (while respecting maximal and minimal allowable area). When energy must be maximized, it is therefore necessary to keep cross-sectional area minimal in this region. Because of the symmetry of the optimal tract, the area up to the same distance from the open end must be maximal.

In the transition between the narrow section and the wide section, area must be decreased in parts with positive radiation pressure and increased in parts with negative radiation pressure (illustrated in the middle part of Fig. 3). However, decreasing area in a part with positive radiation pressure decreases the radiation pressure, while increasing

---

**Fig. 3:** Tracts for maximizing the first resonance. It is shown how a gradual transition can maximize energy, and thus resonance frequency.
area in a part with negative radiation pressure increases radiation pressure. This occurs because pressure \( p \) and volume velocity \( u \) must be equal in adjoining sections. Because \( u = vA \), decreasing cross sectional area \( A \) of a narrow slice of the tract increases particle velocity \( v \). As pressure remains constant, it follows from Eq. (8) that radiation pressure must decrease. Similarly, increasing area increases radiation pressure. The relation between energy, radiation pressure and area leads to an energy maximum in the transition zone where radiation pressure is zero.

To investigate the exact shape of the transition zone, it is convenient to express the standing wave in the acoustic tract as the superposition of two waves traveling in opposite directions. The particle velocity \( v \) at a given position can now be expressed as \( v = v_0 + v_\pm \). If pressure \( p \) is set to \( p = p_0 + p_\mp \). Both quantities depend on time and place.

For traveling waves in an acoustic horn, the relation between the impedance of the forward traveling wave and the backward traveling wave is as follows:

\[
z_\pm = -z_\mp^*.
\]

where \( z_\pm \) and \( z_\mp^* \) are the forward and backward impedances, respectively and \( z_\mp^* \) indicates the complex conjugate. This follows from the definition of impedance and from Eq. (1).

Because radiation pressure must be zero, and using the definition of impedance, it follows from (8) that:

\[
\frac{\left(z_\pm v_0 + z_\mp^* v_\pm\right)}{\rho c^2} = \frac{\rho (v_0 + v_\pm)}{2} = \frac{\rho (v_\mp + v_\pm)}{2}.
\]

which, by using (20) and by dividing impedance by \( \rho c \), can be written as:

\[
\frac{\rho \left(z_\pm v_0 + z_\mp^* v_\pm\right)}{\rho c} = \frac{\rho (v_\mp + v_\pm)}{2}.
\]

which is only possible if:

\[
\frac{z_\pm}{\rho c} = \frac{z_\mp^*}{\rho c} = \pm i
\]

The sign of \( i \) only determines which of the two waves is called the forward wave and which the backward wave. For convenience the positive value is chosen. The relation between pressure and particle velocity then becomes:

\[
\frac{p}{\rho c} = i v
\]

The equations for pressure and volume velocity (1) can be rewritten for pressure and particle velocity as follows:

\[
\frac{dA}{dx} v + A \frac{dv}{dx} = -A \frac{\partial p}{\rho c^2} \frac{\partial t}{dx}
\]

\[
\frac{\partial p}{\partial x} = -\rho \frac{\partial v}{\partial t}
\]

substituting (24) gives:

\[
\frac{dA}{dx} v + A \frac{dv}{dx} = \frac{A}{c} \frac{dv}{dt}
\]

\[
\frac{ic}{\partial x} = \frac{dv}{dt}
\]

Combining these two equations with some simplification gives:

\[
\frac{dA}{dx} v = -2i A \frac{dv}{c \partial t}
\]

Now, because it is investigated what the resonance frequency of the tract is, it can be assumed that oscillations are harmonic, and that therefore \( \frac{dv}{dt} = i \omega \cdot v \). Simplifying (27) with this, dividing out \( v \), and using the definition of \( k \), gives:

\[
\frac{dA}{dx} = 2kA
\]

which solves to:

\[
A = A_0 e^{2i k x}
\]

where \( A_0 \) is a constant that gives the area at the throat of the horn. This means that the transition zone consists of an exponential horn that expands at a ratio of two times the wave number \( k \). This means that the horn is operating at its cutoff frequency [e. g. 15, sections 198 and 201].

The final question is what the exact value of \( k \) is for a tract of given maximal and minimal diameter. For maximizing energy, the starting point of the horn is exactly the point where radiation pressure first becomes zero. As the tract up until that point has constant area, the solution for the wave equation in a tract of constant area can be used. In the case of a tube with constant area that is closed at one end, radiation pressure becomes zero at a length \( l_c \) such that \( kl_c = \pi / 4 \). For reasons of symmetry, the same must be true for the cylindrical part of the tract near the open end. Furthermore, given (29) and given the minimal and maximal areas, the length of the horn section can be calculated to be:

\[
l_c = \frac{\ln A_{\text{max}}/A_{\text{min}}}{2k}
\]

Now the total length of the tract \( L \) is equal to \( 2l_c + l_e \). Solving for \( k \) it is found that:

\[
k = \frac{\pi + \ln A_{\text{max}}/A_{\text{min}}}{2L}
\]

The optimal tract is illustrated in the right part of Fig. 3. This concludes the analysis of the tracts with maximal and minimal first resonances.

### 2.3 Higher resonances

From the optimal tracts for the first resonance (which will be called basic tracts for short), optimal tracts for higher resonances can be constructed straightforwardly. The basic tracts are the shortest (for minimal resonance frequency) or longest (for maximal resonance frequency) that connect a pressure node (at the closed end) to an anti-node (at the

![Fig 4: derivation of tracts with minimal and maximal 2nd resonances from tracts with optimal 1st resonances.](image)
open end). Connecting three basic tracts for the first resonance front-to-front and back-to-back (as illustrated in Fig. 4) must therefore result in the shortest or longest possible tract with that frequency as the second resonance. If there were a differently shaped tract that had a lower (higher) resonance, then there would have to be a subsection of that tract in which a pressure node is connected to an anti-node that is shorter (longer) than the basic tract. This is in contradiction with the assumption that the basic tract is the shortest (longest) possible tract for that frequency. Hence the constructed tract must be optimal. Tracts for even higher resonances can be generated analogously by adding two basic tracts per resonance.

3 Conclusion and Discussion

The exact shapes for acoustic tracts minimizing and maximizing resonance frequencies have been derived. The shape for the tract with the minimal resonance frequency is a tube with constant maximal area near the closed end, followed by a tube of equal length and constant minimal area. The shape for the tract with maximal resonance frequency consists of a tube with constant minimal area connected to an equally long tube of maximal area through an exponential horn that has the resonance frequency as its cutoff frequency. The wave number of the minimal resonance is given by Eq. (13) and that of the maximal resonance is given by Eq. (31). Tracts for higher resonances can easily be derived from these tracts by concatenating multiple copies, as shown in figure 4.

The analytical tools – symmetry of the wave equation and the use of adiabatic invariance, can possibly be used for other problems as well, such as finding the vocal tract shape in which two resonances are maximized or minimized simultaneously, although this would involve more complicated mathematics.

An interesting observation is that for maximizing resonance frequencies, gradual transitions are necessary, while for minimizing resonance frequencies, abrupt transitions are necessary. It appears that this observation has never been made before in the literature, and it supplements (and should not be confused with) Fant’s classical observation that “The general rule is that a reduction of area contrasts within the vocal tract shifts F-pattern towards a neutral vowel” [6, p. 81].

As real vocal tracts consist of flexible, continuous tissue, gradual transitions occur more readily than abrupt transitions, and one would expect biological signals to be closer to the theoretically maximal resonance frequencies than to the theoretical minimal ones. Also, relations (13) and (31) can be used to estimate a range for the ratio between minimal and maximal cross-sectional area from measurements of resonance peaks of a signal and given the length of the vocal tract that produced it (which is usually easier to measure). This can be done by substituting the observed length and wave number, and solving for $A_{\text{min}}/A_{\text{max}}$.

Acknowledgements

This work is part of the NWO vidi project “Modeling the evolution of speech”. The author wishes to thank Stephan de Rooide for useful discussion and comments.

References