On Formants

R. S. McGowan

CReSS
Books

On Formants

Printed in the United States of America

First printing 2020

ISBN: 978-0-9997574-2-0

CReSS Books
1 Seaborn Place
Lexington, MA 02420

www.cressbooks.com

for my teachers,

Louis Goldstein, Michael Howe, and Philip Rubin

Table of Contents

Chapters and sections marked with an * may be omitted on a first reading.

Preface

Formants are "...one of several characteristic bands of resonance, a combination of which determines the distinctive sound quality of a vowel..." according to the *Oxford English Dictionary*. The first use of the term formant was in an 1894 German publication, and it first appeared in English in a 1901 issue of *Nature*. The term has been used to refer to any spectral prominence, but we restrict ourselves to resonances of the vocal tract. Also we recognize that speech segments other than vowels are, at least partly, characterized by their formants. Because formants and resonances are taken to be synonymous here, each formant is associated with a frequency, its formant frequency, which is simply a resonance frequency.

In fact, the magnitudes of raising and lowering of formant frequencies from their values, say, for a neutral vocal tract shape, are very important ways that sounds of speech are characterized. In this book, we consider the causal relation between cross-sectional areas of a tube and the first two formant frequencies. Tubes are employed to explain the acoustic properties of a vocal tract. Cross-sectional areas are the areas of the tube in planes perpendicular to the tube axis. If cross-sectional areas are associated with positions along the tube axis, we obtain an area function. Understanding the relationship between tube area functions and formants is an intermediate step toward understanding the causal relationship between articulation and acoustics.

Among the findings presented in this book are 1) the axial length of a constriction has a large effect on formant frequencies, 2) that a mathematical object, spatial phase, is an important concept in understanding how formant frequencies change with changes in tube cross-sectional area, and 3) it appears possible to characterize monophthong vowel production in an abstract form of area function parameters. Chapters and sections that may be omitted on a first reading of this book are marked with an * in the Table of Contents.

The findings in the present book rely on the results published in *Acoustics of Speech Production* (McGowan 2018). A short introductory chapter may enable the reader to understand the present book, however a more compete understanding of the results presented here can be obtained by reading McGowan (2018).

I have many people to thank for their consistent encouragement regarding this work. These people include Richard Goldhor, Joel MacAuslan, Terry McKiernan, Susan Nittrouer, and Mark Tiede. This book is dedicated to my three most important teachers in phonetics, fluid mechanics and acoustics, and writing: Louis Goldstein, Michael Howe, and Philip Rubin.

Richard McGowan
Lexington, Massachusetts

Chapter 0: Introduction

This book is about the relationship between vocal tract shape and speech acoustics in terms of formants. As in many scientific areas, there was substantial effort in speech acoustics post World War II, notably with the work of Gunnar Fant and Ken Stevens. They modeled speech acoustics by employing electrical analogues, and the field saw remarkable progress with constantly growing understanding of the relationship between vocal tract shape and the acoustic result for a listener. The work emanated from laboratories in Stockholm, Sweden; Cambridge, Massachusetts; Murray Hill, New Jersey, as well as many places around the world including those in England, France, India, Japan, and others. We cannot name all the individuals who have contributed to the knowledge base in speech acoustics, and many people are currently adding to that base.

While the basic mathematical modeling and computational work in speech acoustics has used the parlance of electrical analogues, other aspects of the related physics of air flow and tissue vibration have been successfully understood to a good degree. These include sound production at the glottis and fricative noise production, which require that we go back to basic physical mechanisms not related to electrical analogues. Again, there are many scientists who have made contributions to these areas, and this group often overlaps with the group who have studied speech acoustics. Many of these researchers are mechanical engineers and physicists who have worked in mathematical modeling of physical systems and use the language of classical mechanics. The present author is one of these people; he has been fortunate to be a part of the research effort into the physics of speech.

The purpose of this book is to provide a working knowledge of, and intuition in, speech acoustics for people who do not work directly in this subject, but who have an interest in speech production and phonetics. The book treats speech acoustics as a part of classical mechanics. The author is most comfortable in this milieu and the acoustic and non-acoustic aspects of air motion, as well as air-tissue interaction, can be discussed within a unified mathematical theory in later applications of the findings presented in this book.

The mechanics of acoustic air motion in a tube is described in a *field theory*, where each point in the space represents a very small volume of air with many molecules, and where the particular volume of air at a spatial location can be different from moment-to-moment (McGowan, 2018; Chap. 1). Each point of space is assumed to possess a pressure p and a particle velocity u. Pressure p actually denotes a small volume's pressure difference from a constant ambient pressure, and the particle velocity u is the velocity of the small volume. We most often assume one-dimensional acoustic motion along the axis of a tube, so that we need only to be concerned with the one spatial dimension along the axis of the tube. Therefore, all quantities of interest are assumed to be constant in planes perpendicular to the tube axis. Let x denote the position along the axis of the tube, so that $p = p(x, t)$ and $u = u(x, t)$, where u is the velocity in the x-direction and t is time. Instead of u, it is most often convenient to use volume velocity $Q = A \times u$, where $A = A(x)$ is the cross-sectional area of the tube at position x. This definition makes sense, because u is the same throughout the cross-sectional area under the assumption of one-dimensional motion. In physics we consider the acoustical problem for the tube to be solved when $p(x, t)$ and $Q(x, t)$ or $u(x, t)$ are known for all x in the tube for all time t.

This author has come to believe that classical mechanics is the best way to present speech acoustics to interested people. There is a brief initial steep learning curve in taking the mechanical route. However, eventually one realizes that electrical analogues provide further abstractions from the already abstract picture provided by field theory: electrical analogue models do not provide a good entry into understanding the physics of acoustic wave motion in the vocal tract. Electrical analogues are fine for computational studies of speech acoustics, but by using them we can lead ourselves astray conceptually if we are not careful. Suppose one has derived the electrical analogue appropriate for a particular vocal tract shape. We have essentially derived an electrical circuit, and electrical circuits depend on the spatial ordering of its electrical components, but not on the spatial distance between these components. Thus, unless we specifically ask about how the circuit relates to the spatial distribution of voltage and current, which correspond to pressure p and volume velocity Q, respectively, we simply use the circuit for the computation of quantities such as resonance, or formant, frequencies based on the spatial ordering of electrical elements. We could use the circuit as an analogue computer to provide the spatial distributions of voltage and current if we pay attention to the spatial distance, as well as the ordering, of the electrical components. This step often is not taken, and the ideas of spatial distance and spatial distribution of physical quantities are neglected.

In *Acoustics of Speech Production* (McGowan, 2018; Chaps. 9 & 10), we noted the importance of satisfying certain *boundary conditions* on pressure p and/or volume velocity Q at either end of the vocal tract tube in determining formant frequencies. In the simplest model we require that $Q = 0$ at the glottal end and that $p = 0$ at the mouth end in order to obtain a resonance, or formant. In a tube of variable cross-sectional area, $A = A(x)$, we assume that p and Q have sinusoidal time variation at frequency F, or circular frequency $\omega = 2\pi F$. Instead of calculating p and Q, we calculate their relative magnitudes and phases. The relative amplitudes and phases can be expressed in terms of a complex number ratio of Q to p known as an *effective admittance*. Given the boundary values for p and Q, we can state the boundary values for the effective admittance: at the glottal end effective admittance is 0, because $Q = 0$ there, and at the mouth end effective admittance is ∞, because $p = 0$ there. Starting at either the glottal end or the mouth end, an iterative calculation of effective admittance is initiated toward the other end of the tube by taking small steps in x. As $A(x)$ varies, the assumption of continuity of p and Q as a functions of x, and therefore continuity of effective admittance in x, enables these iterative calculations. At the terminating end, another boundary condition applies to the effective admittance. It turns out that the effective admittance meets the boundary condition at the terminating end only for certain frequencies F_n for $n = 1, 2, 3, \ldots$, which, of course, are the formant frequencies. This is essentially the same calculation that is done with electrical analogues, in order to derive formant frequencies. We can use the effective admittance calculations to compute the spatial distributions of pressure p and Q throughout the vocal tract tube, just as with an analogue computer based on an electrical circuit. In most cases, we do not make the step of calculating spatial distributions, because we are concerned primarily with formant frequency, which is determined by the terminating boundary condition.

In this book the procedure is reversed, and the spatial distributions of acoustic quantities p, Q, and u are treated as the primary quantities of interest. We can use these spatial distributions of pressure, volume velocity, and particle velocity throughout the vocal tract tube to inform us about formants and their frequencies. We use this strategy of examining spatial distributions of acoustic quantities in order to increase our understanding of formants and on how formant frequencies are raised and lowered. Instead of the iterative procedure of calculating effective admittance from one end of the tube to the other, we obtain spatial distributions of p, Q, and u using the concept of *standing waves* (McGowan 2018; Chap. 5). We examine the spatial distributions of acoustic quantities p, Q, and u that correspond to formants, which are represented by standing waves. We also examine the spatial distributions of other quantities derived from the spatial distributions of p and u, to discover how formants and their frequencies behave with changes in vocal tract tube shape. These quantities are discussed below.

From many sources, including McGowan (2018), we know that one-dimensional acoustic motion for tubes closed at one end and open at the other can be represented as standing waves for p, Q, and u [1]. Let $f(x,t)$ represent either p, Q, or u, then throughout this book we assume a standing wave representation,

$$f(x,t) = f^{space}(x) f^{time}(t) \tag{1}$$

It appears that the standing wave representation means that we separate the spatial variation from the time variation of the acoustic quantity. This appearance is completely deceptive. Standing waves can be written as sums of traveling waves (McGowan, 2018; Chap. 5). Let c_0 be the (adiabatic) speed of sound. For quantities that satisfy the *wave equation*, such as p, Q, and u, values remain constant as long as $x - c_0 t$ for right-going traveling waves remains constant, or as long as $x + c_0 t$ remains constant for left-going traveling waves

[1]Mathematically, an infinite sum of standing waves can include the portion of acoustic energy that leaves the opening. However, it may be practically best to think of a standing wave along with a small of a traveling wave that accounts for acoustic loss through the opening. We only need to consider the standing wave to understand formants.

(McGowan, 2018; Chap. 2). Thus, space x is intimately related to time t through the parameter c_0, In fact, the formant frequencies are not directly determined by the function $f^{time}(t)$, but by the function $f^{space}(x)$. If $\omega = 2\pi F$ is a circular frequency that appears in $f^{time}(t)$, then the parameter ω/c_0 appears in $f^{space}(x)$. Because of the fact that two boundary conditions, one at each end of the tube, must be satisfied, ω/c_0 is restricted to be one of the special values $(\omega/c_0)_n$ for $n = 1, 2, 3, \ldots$ Of course this means that ω is restricted to ω_n, and that $F = \omega/(2\pi)$ is restricted to $F_n = \omega_n/(2\pi)$. These are the formant frequencies, and the spatial distributions $f^{space}(x)$ are required to derive these. Often we can set $\omega/c_0 = 2\pi/\lambda$, where λ is wavelength, so that determining the number of wavelengths of *signed amplitudes* $p^{space}(x)$, $Q^{space}(x)$, and $u^{space}(x)$ determines formant frequencies. It seems that these signed amplitudes are spatial distributions that we need. Later in the book, we replace this procedure of counting wavelengths with the concept of *spatial phase.*

As stated initially, this book is about the relationship between vocal tract shape and speech acoustics. Peter Ladefoged remarked that the traditional terminology of tongue height, tongue fronting, and lip rounding corresponds more closely to an auditory or an acoustic descriptions of vowels in terms of formant frequencies than the articulatory-like names connote (Ladefoged, 1993; pp. 78-80). It appears that we do not consistently relate the details of articulation, vocal-tract geometry, and formant frequencies. Phonetics needs more ready knowledge of, and intuition in, this relationship to help in the project of putting vowel production on an articulatory footing. The lack of knowledge of the causal relationships has meant that a truly articulatory phonetics that includes vowel production has been hindered. In fact we can include all *sonorants* along with the vowels as a class of segments where more knowledge is needed.[2]

In Chapter 1 we begin this book with a review of concepts and results, for which the reader may want to consult McGowan (2018) for more detail. We review the idea of resonance, mode, or formant along with the results for a tube of constant cross-sectional area,

[2]Sonorants, such as [ɹ], are speech segments where the upper vocal tract contains neither constrictions that completely block air flow, nor constrictions that produce significant turbulence noise. The other speech segments are called *obstruents* (Ladefoged, 1993; 62-3).

known here as a *uniform tube*. We state the well-known result that p and Q, and, therefore u, have signed amplitudes, $p^{space}(x)$, $Q^{space}(x)$, and $u^{space}(x)$, that are sinusoidal in the spatial variable x. (See the discussion around Equation (1) for the notion of signed amplitude.) Sinusoidal variation in space is important because the concept of wavelength λ is well-defined in such situations. As suggested above, we want to be able to count number of wavelengths within a tube, and one of the ways to do this within a tube with variable cross-section is to approximate the tube with a set of uniform sub-tubes. We simply add the counts of wavelength in each sub-tube for a total number of wavelengths within the overall tube to determine formant frequency. We go on to define time-averaged *kinetic energy density* and *potential energy density*. These quantities are a function of the spatial variable x like the signed amplitudes, and they can be computed from $u^{space}(x)$ and $p^{space}(x)$. *Steady pressure* is stated to be the difference between these energy densities, and this quantity is important for *acoustic perturbation theory*. This theory is reviewed in Chapter 1, followed by a review of the idea of a Helmholtz resonator.

Chapter 2 examines area function changes that are local, which means that they are short in the axial direction. We use acoustic perturbation theory to explain the patterns observed as short constriction degrees are varied in different portions of a tube. We do not use the signed amplitudes directly in this investigation, but only indirectly through the steady pressure's dependence on the signed amplitudes.

In Chapter 3 we study the second formant in detail as constriction location, constriction degree, and constriction axial length vary. Signed amplitudes for p, Q, and u, as well as energy densities, and steady pressure are examined and compared with these quantities for the tube of constant cross-sectional in order to understand how the formant frequency changes occur. The relationship between wavelength and phase jumps between uniform sub-tubes explains changes in wavelength, and, therefore, these phase jumps help determine formant frequency. As a result of its importance in the physics of formants, we define a mathematical object called *spatial phase*, which contains information about phase jumps and is itself a spatial distribution.

Chapter 4 develops the idea of spatial phase further. Its centrality in understanding the relationship between vocal tract shape and formants is argued. More mathematical and computational aspects of spatial phase are explored in the latter part of the chapter, and this part may be omitted on a first reading of this book.

Chapter 5 contains a detailed study of short constrictions and expansions. This chapter may also be omitted on a first reading of this book.

In Chapter 6, we apply some of the knowledge gained in the rest of the book to derive area functions for eight American English vowels from their first three formant frequencies. Deriving vocal tract shape from acoustic data is an old and venerable problem in speech production (*e.g.* Mermelstein, 1967; Schroeder, 1967; Schroeter & Sondhi, 1994; Story, 2006; Kaburagi, 2014). Because of the limited acoustic data in our case, we proceed cautiously by deriving what may be termed abstract vocal tract shapes. However, we believe that this is a step toward building an articulatory description of vowel production.