Physical and Linguistic Aspects of Speech

R. S. McGowan

CReSS
Books

Physical and Linguistic Aspects of Speech

First printing 2024

ISBN: 978-0-9997574-0-6

CReSS Books
1 Seaborn Place
Lexington, MA 02420

cressbooks.com

for Michael Howe,

a friend, a colleague, and an excellent teacher

It is not in the heavens … But the word is very close to you, in your mouth and in your heart to do it.

Deuteronomy 30:12, 14 (translated by Robert Alter)

You've always had the Power my dear, you just had to learn it yourself.

Glinda in the movie entitled "The Wizard of Oz"

Table of Contents

Physical Themes

Theme II: Formants

Theme IV: Sources (and Sinks) of Acoustic Motion

Themes Interlude on the Mechanics of Air Motion

Theme V: Non-Acoustic Air Motion Can Be a Source for Acoustic Motion

Theme VI: Tissue Vibration

Linguistic Examples

Example I: Young Children's Productions of [ɹ], [u], and [ɑ]

Example II: A Five-Parameter Area Function Vowel Space

Example III: Sensitivity of Monophthongs in the Five-Parameter Area Function

Example IV: Modified Auditory Feedback

Example V: Hyper-Articulation

Example VI: Acoustic Sensitivity and Peripherality of American English Monophthongs

Example VII: Monophthong Tenseness in American English

Example VIII: Source-Tract Interaction

Example IX: Development of Velar and Alveolar Stop Releases

Example X: Preliminaries to an Analysis of Weak Fricatives

Example XI: Designing Inverse Mappings for Speech Production

Example XII: A View of Speech Perception

Appendix B: Spatial Phase and Pressure Amplitude

Appendix C: Parameter Sensitivity with the Five-Parameter Area Function

Appendix D: A recasting of and errata for: McGowan, R.S. & Howe, M.S. (2012). "Source-tract interaction with prescribed vocal fold motion." *Journal of the Acoustical Society of America*, **131**, p 2999, **and the connection between the source-tract and aeroacoustic pictures of the voice source.**

Appendix F: Acoustic Perturbation Theory and Parameters

Preface

The basic idea that I write a book that summarized my previous two books and to entertain issues in speech production was Philip Rubin's. I simply needed to structure and write it. Fortunately, I find writing to be psychically rewarding. The general idea was to take the results of my previous two books: *Acoustics of Speech Production* and *On Formants* and present the results largely without mathematical justification in the main body of the text. I hope to have done this in such a way as to make the results of mathematical physics compelling for the non-physicist, particularly for speech clinicians, linguists, and psychologists who are involved in the study of speech production. (And don't worry, there are some physicists who do not get along well with the mathematical side of physics, as is the case in all other fields.)

The book is divided into Physical Themes and applications of these Themes in Linguistic Examples. Of course there is not a strict division of physics and linguistics between the Themes and Examples, because the main thrust of the book is to show that a good understanding of physics can move phonetics and, perhaps, phonology forward. Mathematical details, particularly for Theme II and Example III, are contained in the Appendices. These Appendices also contain clarification and errata for a published paper, as well as a mathematical summary of another published paper. More details on what to expect in the book are discussed in the Introduction below.

A host of people helped make this book possible. I thank Louis Goldstein and Jeff Mielke for conversations on linguistic sound change, which I consider in Example VII. Both people recommended the classic book by William Labov on the internal factors for linguistic sound change. Louis Goldstein encouraged my research on vowel tenseness based on a hunch I had on how to characterize vowel tenseness in a phonetic way, as well as recommending that I read the important work of Sidney Wood from the 1970s. I thank Andy Wedel for a conversation that motivated me to formulate Theme III as a sequel to Theme II. He also provided me with helpful references on hyper-articulation, which led me to include Example V. Richard Wright and I had some email exchanges and he supplied references that were critical for me in understanding hyper-articulation research. Robert Remez

supplied references and friendly, well-written emails on his views on speech perception. Louise Ratko and her colleagues, Felicity Cox and Michael Proctor, provided invaluable insight into the current phonetic understanding of Australian English, as well as its evolution. There are a number of people with whom I communicated and who supplied references in their areas of expertise. These include: Diana Archangeli, Rob Hagiwara, Khalil Iskarous, Carol Fowler, Phil Hoole, Caroline Huang, Michael Krane, Ewen MacDonald, Tine Mooshammer, David Purcell, Enn Rosen, Mark Tiede, and Sidney Wood.

People who read a version or a part of this book and offered comments were Michael Howe, Michel Jackson, Robert Remez, Philip Rubin, and Richard Wright. Philip Rubin provided detailed comments throughout a complete version of the present book.

I thank John A. Johnson, a fine physics professor at Kenyon College in the early-to-mid 1970s, who had me read James Lighthill's book on generalized functions (Lighthill 1962), as well as study the proof of the Riemann mapping theorem as part of an independent study. I also thank the mathematics faculty of Kenyon College during the same period: Chris Duckenfield, Robert Fesq, Daniel Finkbeiner, Wendel Lindstrom, Robert McLeod, and Stephen Slack for providing me with a well-grounded education in mathematics. Helene Shapiro and I were part of a small group of students of mathematics at Kenyon College in the 1970s. When I recently asked her for her favorite book on Hilbert spaces and described my problem of trying to prove a particular set of eigenfunctions to be complete, she suggested a productive fall-back position that led me to discover some recent mathematical work on Green's functions on direct sums of Hilbert spaces, which is mentioned in Appendix G.

Richard McGowan
Lexington, Massachusetts

Introduction

Speaking the word “start” involves coordinated tongue, jaw, larynx, and lip movements. The following is a description of these movements along with the consequences of these movements for air motion, including acoustic air motion, in the vocal tract. The tongue tip forms a tight constriction with the palate, while the glottis is more or less open so that the lungs supply air flow during the [s]. The air flow downstream of the tongue-tip constriction is mostly a turbulent jet. The turbulence produces pressure fluctuations at the teeth, which creates noisy sound (i.e. random acoustic motion). This is followed by a brief closure of the vocal tract with the tongue tip in contact with the palate for [t]. During this contact the acoustic motion has decayed, and the vocal folds have attained a more approximated posture in preparation for voicing after the [t] closure has been released. Also, perhaps before noise production has ended for [s], the body of the tongue has started moving into a low, back position appropriate for the upcoming production of the vowel [ɑ]. After the tongue closure for [t] is released, a sufficient glottal air flow is attained very rapidly for voicing with approximated vocal folds, which help make the *formants* of the changing vocal tract salient as they approach those for [ɑ].[1] During the production of [ɑ], the articulators ready for the voiced rhotic [ɹ]. This means that the lips are beginning to round, a tongue constriction at the mid-palate is being formed, and, perhaps, the tongue constriction in the pharynx is moving higher. This happens relatively gradually – on the time scale that is a substantial fraction of the entire voiced segment that includes [ɑ] and [ɹ]. A gliding from one segment to the other is heard by a listener. The tongue-tip returns to the alveolar ridge to close off air flow, including the acoustic air motion, for the final [t] segment.

I believe that a complete physical description of the production of the word “start” includes both articulation and the motion of air, including the acoustic aspect of air motion (Joos 1948). Students of speech production should have some knowledge of the relationships between articulation and air motion, including both acoustic and

---

[1] Formants are understood to be the resonances of the vocal tract.

non-acoustic air motion. Why should anyone be interested in acoustic-articulatory relationships of speech production? If one is interested in acoustic-articulatory relationships, how should they approach the study of these relationships?

This first answer to "why" is for physicists. Speech production involves a system that has not been adequately described according to current physical knowledge. This is nearly enough reason for a physicist. The phrase "nearly enough reason" is used because physicists don't want to waste time describing just any system. The vocal system is very important for our species. A physical description of speech production would seem to be, at least, on par in importance with descriptions of animal gait, flight, and swimming. However, the pure physics of a situation does not provide a good reason for, say, a linguist, a speech therapist, or an experimental psychologist to be concerned with articulatory-acoustic relationships. On the other hand, a physicist is welcome to study the physics of the relation between, say, articulation and acoustics all they want; how does this physical knowledge inform other people's particular questions?

Physicists are concerned with what are known as *forward problems*, which, for speech, I take to be the causal relations that determine acoustics from articulatory configurations in the case of speech production. Physicists are also very interested in *inverse problems*, as exemplified by all the activity at particle colliders around the world. The experiments with these colliders provide data that permit theorists to build models of the ultimate constituents of the universe. The theorists answer the question: What are the particles/fields that would give us the pattern seen in this particular particle scattering experiment? Based on the hypothesized particle/field construct, these theorists predict certain patterns in future experiments. Inferring articulation from speech acoustics is also an example of an inverse problem. And, speech data are much cheaper to obtain than are collider data. Often having good knowledge of the forward problem makes solving the inverse problem easier, or even, practicable.

It is clear that speech clinicians ought to understand the relationships between articulation and acoustics because they are most often charged with helping people alter their articulatory behavior. In many practical situations, a therapist has only his or her own auditory system, but they may also have the means to perform instrumental acoustic and air flow analyses. They must infer the articulation of clients and offer advice to the client on how to change articulatory behavior – they solve inverse problems constantly. Physical knowledge of articulatory-acoustic relationships can only enable better solutions of the inverse problems that confront the clinician.

The motivations for the linguist and the experimental psychologist who are concerned with speech production to know more about articulatory-acoustic relationships are, perhaps, less straight-forward. Part of the reason for this book is to provide motivation for these people beyond the fact that it is knowledge connecting two aspects of speech production. I use the term *speech scientist* to refer to those concerned with the clinical aspects of speech production, as well as linguists, experimental psychologists, and physicists who are concerned with speech production.

As for how we should get to know articulatory-acoustic relations, I make the claim that these relationships should be considered partly under the auspices of physics. Using knowledge of classical physics means employing about 300 years of substantial understanding of the causal relations between vibrating objects and sound waves that reach our auditory system. The science of fluid motion, of which acoustic air motion is an example, has developed since the time of Isaac Newton (1642-1727). There is Daniel Bernoulli (1700-1782), who's Bernoulli's principle provides a relationship between pressure and fluid velocity. This is a statement of conservation of energy under special conditions of fluid motion. Leonhard Euler (1707-1783) wrote down equations for conservation of mass, momentum, and energy for fluids without viscosity known as the Euler equations. Jean le Rond d'Alembert (1717-1783) discovered the *wave equation*, which describes acoustic air motion. There are numerous mathematical physicists of the 19th and 20th centuries who have advanced the science of fluid mechanics and acoustics, including Lord Rayleigh and James Lighthill.

Another reason that classical physics should be used is that the standard understanding of acoustics in speech production, *source-filter theory*, is rooted in classical physics. In particular, Part I of Fant's *Acoustic Theory of Speech Production'* (Fant 1960), as well as the first three chapters of Stevens' *Acoustic Phonetics* (Stevens 1998) are fundamental to our current understanding of speech acoustics. These authors used the language of electrical analogues, knowing fully that these analogies provide a model for the underlying physics of acoustic air motion.

I have worked on the "how" of speech acoustics and the articulatory-acoustic relationship in two previous books: *Acoustics of Speech Production* (McGowan 2018) and *On Formants* (McGowan 2020). Part of the reason for writing these books was to present the acoustics of speech production in the language of classical mechanics, which lends itself to different emphases from those of the electrical analogues. One important consequence of using classical mechanics is that acoustic theory can be *embedded* into the more general theory of the mechanics of air motion, that is, fluid mechanics.

By making the embedding of acoustics within fluid mechanics explicit, the physical appropriateness and accuracy of speech acoustics theory for air motion during speech production can be judged. Many speech scientists already understand that an acoustic description is not appropriate for most of the air motion near a constriction, say during fricative production. However, it should be known more precisely how the pressure fluctuations caused by air motion described using a non-acoustic theory interact with air motion described by acoustic theory. These two types of air motion must interact to some extent because the physical quantities, such as pressure fluctuations, are common to theories that describe either non-acoustic air motion and acoustic air motion. It is simply that often one theoretic description of pressure fluctuation is more appropriate than another.[2] In fact, when I write "acoustic air motion", this should be understood to be air motion described by what is called the *acoustic approximation*, which is discussed in Theme I. By embedding acoustic theory within the larger context of air motion in general, we can understand how air motions described by a non-acoustic theory can provide sources for acoustic motion. This has important consequences in the understanding of the physics of speech production.

Instead of articulation, I work with slightly more abstract objects, or functions, called *area functions* in this book. I conceive of a vocal tract for which a mid-line from the glottis to the opening at the lips has been defined along with planes perpendicular to this mid-line. These planes have well-defined areas within the vocal tract, which

---

[2] Theorists take advantage of the fact that sometimes there are spatial regions where both descriptions are approximately valid.

are called *cross-sectional areas*. An equivalent tube with a straight mid-line is defined to have the same cross-sectional areas as a function of position along the mid-line as those of the vocal tract. An example of such a tube is shown in Figure In-1, where the mid-line runs along the x-axis. Given a tube, the area function $A(x)$ is the relation between axis position x to the cross-sectional area A at the position x along the mid-line. The area function for the tube in Figure In-1 is represented graphically in Figure In-2. Numerical values for the area are not given in this figure, and the open disc and filled disc at $x = 0$ indicates that there is a hard wall perpendicular to the x-axis at $x = 0$. [The open disc at $x = 0$ simply indicates that the area function at $x = 0$ is given by the closed disc.] Area functions determine the one-dimensional acoustic propagation along the axis of a tube. In this book I consider articulatory-acoustic relationships only in terms of area function-acoustic relationships, and I leave the difficult-to-obtain relationships between articulation and area functions to others. An important part of future speech production research is to employ vocal tract imaging to map between various states of vocal tract articulators and area functions.

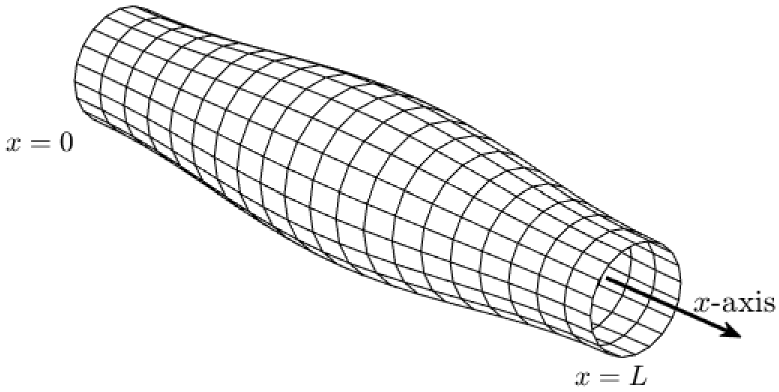

Figure In-1: A tube of length L

Certain aspects of the relationship between articulation and acoustics are already taken into account by the research community. For instance, stop consonants are described largely in terms of place-of-articulation and the acoustic consequences of the stop closure and/or release can be measured instrumentally. The phonetician is trained to detect

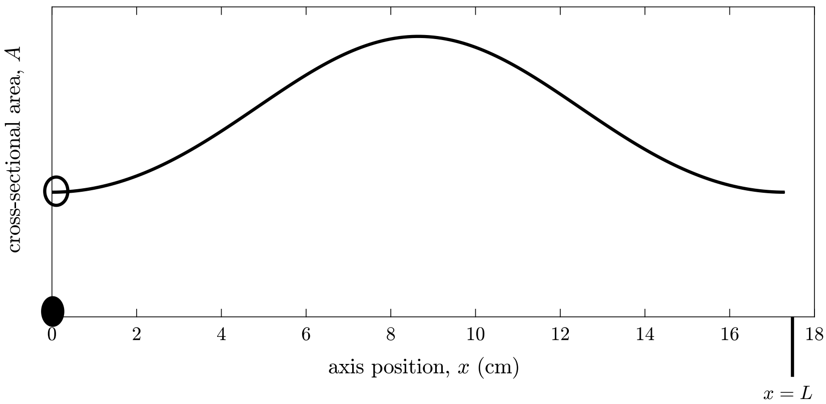

Figure In-2: The area function, $A = A(x)$, for the tube shown in Figure In-1. Values for cross-sectional areas are not specified. The open and filled discs at $x = 0$ indicates zero area, or a hard wall at $x = 0$.

place-of-articulation just by listening. I claim that a speech scientist would benefit with more ready knowledge of the articulatory-acoustic relationship than is currently typical.

Exploring particular articulatory- or area function-acoustic relationships can employ *what-if reasoning* that expands the tool-kit of speech scientists. I have often used this kind of reasoning in my own research. A bit of software can be involved as a part of this what-if reasoning. Based on my own knowledge I am able to write programs that predict both formant frequencies and the relative amplitudes of formants from area functions. That is, I have implemented a software solution to a part of the forward problem. If I am interested in how formant frequencies and relative amplitudes change with a hypothetical change in an area function, I can easily find the answer. With a little more work, and by using known optimization procedures with the software for the forward problem, I am able to infer aspects of area functions that produce certain formant frequencies and relative amplitudes. That is, I am able to perform what-if reasoning for the inverse problem, as well as the forward problem. As important is the fact that I know the approximations that have been made in the process of developing software for the numerical functions.[3] [4] The software for the forward problem can be as crude as a two-tube model, or as complex as a very sophisticated articulatory synthesizer. Also important is the intuition that is instilled in people who know some physical theory of acoustics and use computational methods to explore the relationships between articulation or area functions and the resulting acoustics.

There is an economic motivation for greater knowledge in the relationship between area function and acoustics. Speech scientists are often interested in articulatory aspects of speech, as, for instance, clinicians must be. Obtaining direct articulatory information by imaging using the various technologies of ultrasound, electromagnetic articulography, and magnetic resonance imaging often involves large expense and technical expertise. Acoustic speech data is inexpensive and plentiful. Progress toward solving inverse problems in speech can go a long way to adding knowledge of articulatory behavior in speech.

Current research bases some of its methodology on known articulatory- or area function-acoustic relationships. For instance, in a production and perception study of nasalization in vowel-nasal alveolar consonant sequences at the ends of words, Beddor (2009) inferred the relative timing of tongue closing gestures and nasal port opening gestures using acoustic characteristics found in examples of Fant (1960) and Stevens (1998). Beddor's work directly informs our understanding of historic sound change that results in a nasalized vowel and the loss of the alveolar consonant. While this is one example of excellent work that has used published results on the acoustics of speech production, in general, people need to possess the knowledge and tools to perform what-if reasoning in situations that have not been worked out previously.

The world in which basic knowledge of, and research interest in, speech acoustics residing in an engineering department at a university where there are other speech scientists appears to be ending. At least, the ratio of all speech scientists to those with knowledge of speech

---

[3] Algorithmic optimization can be considered to be rule-based what-if reasoning.

[4] What-if reasoning is related to the abductory reasoning identified by C. S. Peirce, which can be called hypothesis making (Peirce 1955, 1992a). While what-if reasoning can lead to a hypothesis, it is a less committed search of relationships between physical quantities. An hypothesis may arise from the search.

acoustics is becoming large. Contemporary speech technology has little interest in the physics of speech, which means that disciplines that depend somewhat on this knowledge are, or will be, wanting. We cannot simply depend on knowledge that has been archived in publications. I advocate that all academics and other professionals in the speech sciences have some awareness and competence in physical acoustics. This would involve a cultural change within linguistics, speech and hearing, and psychology departments. For this reason, it may not come to pass, or only come to pass in the distant future.

Ultimately language is a human activity that relies on the brain and cognitive processes. However, physical processes impinge on the brain during the perception of language, and the brain directs motor activity within the physical world during language production. The senses for both perception and production of language can involve audition, vision, and touch. Because of this, the brain and cognitive processes are tightly coupled to the physical world. This is the reason that it is important to possess a good grounding in the physics of, say, speech production in the study of human language.

Other than this Introduction, a transition, and an interlude, this book has two different kinds of chapters called Themes and Examples. The Themes are about the general ideas discussed largely in McGowan (2018) and McGowan (2020). Many of the Themes are directly relevant to the source-filter theory of Fant and Stevens. Theme I is about the idea of theory embedding introduced above. Themes II and III are about the filter side of source-filter theory, which may broadly be thought of as a study of possible acoustic motion in a tube. Themes IV and V are about sources of acoustic energy, and, thus, may be considered to be about how observed acoustic motions arise. The final theme, Theme VI, is about tissue vibration that is induced by air motion in the vocal tract. I hope that these Themes orient the reader without a large physical sciences background to what is contained in my previous two books, and how they relate to source-filter theory. Many things in the present book are stated as fact without proof. The reader who wants proof can refer to McGowan (2018, 2020), and/or, to the references listed in those books. The supporting mathematics for new results in this book, as well as other topics, are presented in the Appendices.

There is also a parallel series of Examples from the speech sciences that follow the Themes. Examples are drawn from others' and my own research, and they are intended to highlight the importance in knowledge of articulatory- or area function-acoustic relationships in speech science. I present some relatively advanced fluid mechanics field theory in Example X on weak fricatives. There is no apology for this, as I advocate that some results of mathematical physics be included into speech production and phonetics: it is necessary to make progress in certain areas. Minor examples are provided within Themes.

Despite the lack of mathematical proof, the reader should get a feel for the nature of how results are obtained in mathematical physics, as well as the nature of deduction and speculation based on those results. A mathematical star in this book is Micael Howe's *Theory of Vortex Sound* (Howe 2003), which is applied to understand sources of acoustic motion in the vocal tract. Another mathematical star of the book is something called *spatial phase*, which was introduced in *On Formants* (McGowan 2020). This object is introduced again in Theme II, where mathematical invariants based on spatial phase are also reintroduced. The invariants are called *total change in spatial phase evaluated at formant frequencies, TCSPEFFs*, and they are used extensively this book, as will be seen in Example III and many subsequent Examples. [There is an invariant TCSPEFF for each formant.] For instance, expressions for *formant frequency sensitivity* to changes in area function can be derived using this invariant. In short, the theory of vortex sound and TCSPEFFs are "pure gold" for the purposes of this book. The reader is gradually introduced to the concept of spatial phase in Theme II. The theory of vortex sound is introduced in the Themes Interlude and Theme V.