THIRD EDITION

ACOUSTIC and AUDITORY PHONETICS

KEITH JOHNSON

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Acoustic and Auditory Phonetics

THIRD EDITION

ACOUSTIC and AUDITORY PHONETICS

KEITH JOHNSON

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This third edition first published 2012

Edition history: Basil Blackwell Inc (1e, 1997); Blackwell Publishers Ltd (2e, 2003)

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Johnson, Keith, 1958–
Acoustic and auditory phonetics / Keith Johnson. – 3rd ed.
p. cm.
Previously pub.: 2nd ed., 2003.
Includes bibliographical references and index.
ISBN 978-1-4051-9466-2 (pbk. : alk. paper) 1. Phonetics, Acoustic.
2. Hearing. I. Title.
P221.5.J64 2012
414–dc22

2011009292

A catalogue record for this book is available from the British Library.

This book is published in the following electronic formats: ePDFs [ISBN 9781444343076];
ePub [ISBN 9781444343083]; Mobi [ISBN 9781444343090]

Set in 11/13pt Dante by Graphicraft Limited, Hong Kong

1 2012

Contents

Acknowledgments

I started work on this book in 1993 at the Linguistics Institute in Columbus, Ohio, and am grateful to the Linguistic Society of America and particularly the directors of the 1993 Institute (Brian Joseph, Mike Geis, and Lyle Campbell) for giving me the opportunity to teach that summer. I also appreciate the feedback given to me by students in that course and in subsequent phonetics courses that I taught at Ohio State University.

Peter Ladefoged had much to do with the fact that this book was published (for one thing he introduced me to Philip Carpenter of Blackwell Publishing). I also cherish our conversations about the philosophy of textbook writing and about the relative merits of Anglo-Saxon and Romance words. John Ohala commented extensively on an early draft with characteristic wit and insight, and Janet Pierrehumbert sent me ten long e-mail messages detailing her suggestions for revisions and describing her students' reactions to the manuscript. I appreciate their generosity, and absolve them of responsibility for any remaining errors.

I was very fortunate to work in an incredibly supportive and inspiring environment at Ohio State University. Mary Beckman provided me with encouragement and extensive and very valuable notes on each of the book's chapters. Additionally, Ilse Lehiste, Tsan Huang, Janice Fon, and Matt Makashay gave me comments on the speech perception chapter (I am grateful to Megan Sumner for suggestions and encouragement for the revisions of the speech perception chapter in this edition), and Beth Hume discussed the perception data in chapters 6–9 with me. Osamu Fujimura discussed the acoustic analysis in chapter 9 with me (he doesn't completely agree with the presentation there).

My brother, Kent Johnson, produced the best figures in the book (figures 4.1, 4.5a, and 6.7).

For additional comments and suggestions I thank Suzanne Boyce, Ken deJong, Simon Donnelly, Edward Flemming, Sue Guion, Rob Hagiwara, SunAh Jun, Joyce McDonough, Terrence Nearey, Hansang Park, Bob Port (who shared his DSP notes with me), Dan Silverman, and Richard Wright. Thanks also to those who took the time to complete an "adopter survey" for Blackwell. These anonymous comments were very helpful. Many thanks to Tami Kaplan and Sarah Coleman of Blackwell, who helped me complete the second edition, and Julia Kirk, Anna Oxbury, and Danielle Descoteaux, who helped me complete the third edition.

This book is dedicated to my teachers: Mary Beckman, Rob Fox, Peter Ladefoged, Ilse Lehiste, and David Pisoni.

K. J.

Introduction

This is a short, nontechnical introduction (suitable as a supplement to a general phonetics or speech science text) to four important topics in acoustic phonetics: (1) acoustic properties of major classes of speech sounds, (2) the acoustic theory of speech production, (3) the auditory representation of speech, and (4) speech perception. I wrote the book for students in introductory courses in linguistic phonetics, speech and hearing science, and in those branches of electrical engineering and cognitive psychology that deal with speech.

The first five chapters introduce basic acoustics, the acoustic theory of speech production, digital signal processing, audition, and speech perception. The remaining four chapters survey major classes of speech sounds, reviewing their acoustic attributes, as predicted by the acoustic theory of speech production, their auditory characteristics, and their perceptual attributes. Each chapter ends with a listing of recommended readings, and several homework exercises. The exercises highlight the terms introduced in **bold** in the chapter (and listed in the "sufficient jargon" section), and encourage the reader to apply the concepts introduced in the chapter. Some of the questions serve mainly as review; but many extend to problems or topics not directly addressed in the text. The answers to some of the short-answer questions can be found at the end of the book.

I have also included some covert messages in the text. (1) Sample speech sounds are drawn from a variety of languages and speakers, because the acoustic output of the vocal tract depends only on its size and shape and the aerodynamic noise-producing mechanisms employed. These aspects of speech are determined by anatomy and physiology, so are beyond the reach of cultural or personal habit. (2) This is a book about acoustic *and* auditory phonetics, because standard acoustic analysis tells only partial linguistic truths. The auditory system warps the speech signal in some very interesting ways, and if we want to understand the linguistic

Acoustic and Auditory Phonetics, Third Edition. Keith Johnson.

significance (or lack of it) of speech acoustics, we must pay attention to the auditory system. The linguistic significance of acoustic phonetics is also influenced by cognitive perceptual processing, so each of the chapters in the second half of the book highlights an aspect of speech perception. (3) There are formulas in the book. In fact, some of the exercises at the ends of the chapters require the use of a calculator. This may be a cop-out on my part – the language of mathematics is frequently a lot more elegant than any prose I could think up. In my defense I would say that I use only two basic formulas (for the resonances of tubes that are either closed at both ends or closed at only one end); besides, the really interesting part of acoustic phonetics starts when you get out a calculator. The math in this book (what little there is) is easy. (4) IPA (International Phonetic Association) symbols are used throughout. I have assumed that the reader has at least a passing familiarity with the standard set of symbols used in phonetic transcription.

Semi-related stuff in boxes

There are all sorts of interesting topics on the edges of the main topics of the chapters. So the book digresses occasionally in boxes such as this to informally address selected (greatest hit) questions that my students have asked. The topics range from underwater speech to the perception of anti-formants, covering digital numbers and the aerodynamics of freeways along the way. I included these digressions because there is no question so simple that it shouldn't be asked. You may find that some of the most interesting stuff in the book is in the boxes.

Improvements Made in the Third Edition

Thanks to the many readers, teachers and students, who have provided feedback about how to improve this book. The main changes that teachers will notice are: (1) I reordered the chapters – putting the presentation of the acoustic theory of speech production earlier in the book and also touching on audition and speech perception early. I realize that there is a good argument for putting the audition and speech perception chapters toward the end of the book, and that the order of presentation that I have chosen presents certain complications for the teacher. I hope that the pay-off – being able to collect acoustic, auditory, and perception data on speech sounds together in each of the chapters 6–9 – is adequate compensation for this. (2) The digital signal processing chapter has been updated to be more compatible with currently available hardware and software, and the linear predictive coding analysis section has been reworked. (3) There is a new speech perception chapter that addresses theoretical issues, as well as the practical concerns that dominated the chapter in the second edition. I adopt a particular stance in this chapter, with which some teachers may disagree. But I also tried to open the door for teachers to engage with the book (and with students) in a theoretical debate on this topic. (4) Sections of the chapters introducing the acoustic theory of speech production, and vowel acoustics, have been rewritten to provide a clearer (and more correct) presentation of resonance and standing waves in the vocal tract. (5) The chapter on audition includes a new section on saturation and masking. (6) Many of the spectrograms in the book have been replaced with ones that are easier to interpret than those found in the previous editions. (7) Each chapter now ends with a selection of recommended readings.

Students won't notice any changes between the third edition and the second – unless you are particularly nerdy and look up old editions of textbooks, or unless you are particularly unlucky and had to retake the course after the publication of this edition. As always, my wish for students who use this book is that learning about acoustic phonetics will be more fun and fascinating with the book than it would have been without it.

Part I

Fundamentals

Chapter 1

Basic Acoustics and Acoustic Filters

1.1 The Sensation of Sound

Several types of events in the world produce the sensation of **sound**. Examples include doors slamming, plucking a violin string, wind whistling around a corner, and human speech. All these examples, and any others we could think of, involve movement of some sort. And these movements cause pressure fluctuations in the surrounding air (or some other **acoustic medium**). When pressure fluctuations reach the eardrum, they cause it to move, and the auditory system translates these movements into neural impulses which we experience as sound. Thus, sound is produced when pressure fluctuations impinge upon the eardrum. An **acoustic waveform** is a record of sound-producing pressure fluctuations over time. (Ladefoged, 1996, Fry, 1979, and Stevens, 1999, provide more detailed discussions of the topics covered in this chapter.)

Acoustic medium

Normally the pressure fluctuations that are heard as sound are produced in air, but it is also possible for sound to travel through other acoustic media. So, for instance, when you are swimming under water, it is possible to hear muffled shouts of the people above the water, and to hear noise as you blow bubbles in the water. Similarly, gases other than air can transmit pressure fluctuations that cause sound. For example, when you speak after inhaling helium from a balloon, the sound of your voice travels through the helium, making it sound different from normal. These examples illustrate that sound properties depend to a certain extent on the acoustic medium, on how quickly pressure fluctuations travel through the medium, and how resistant the medium is to such fluctuations.

1.2 The Propagation of Sound

Pressure fluctuations impinging on the eardrum produce the sensation of sound, but sound can travel across relatively long distances. This is because a sound produced at one place sets up a **sound wave** that travels through the acoustic medium. A sound wave is a traveling pressure fluctuation that propagates through any medium that is elastic enough to allow molecules to crowd together and move apart. The wave in a lake after you throw in a stone is an example. The impact of the stone is transmitted over a relatively large distance. The water particles don't travel; the pressure fluctuation does.

A line of people waiting to get into a movie is a useful analogy for a sound wave. When the person at the front of the line moves, a "vacuum" is created between the first person and the next person in the line (the gap between them is increased), so the second person steps forward. Now there is a vacuum between person two and person three, so person three steps forward. Eventually, the last person in the line gets to move; the last person is affected by a movement that occurred at the front of the line, because the pressure fluctuation (the gap in the line) traveled, even though each person in the line moved very little. The analogy is flawed, because in most lines you get to move to the front eventually. For this to be a proper analogy for sound propagation, we would have to imagine that the first person is shoved back into the second person and that this crowding or increase of pressure (like the vacuum) is transmitted down the line.

Figure 1.2 shows a pressure waveform at the location indicated by the asterisk in figure 1.1. The horizontal axis shows the passage of time, the vertical axis the degree of crowdedness (which in a sound wave corresponds to air pressure). At time 3 there is a sudden drop in crowdedness because person two stepped up and left a gap in the line. At time 4 normal crowdedness is restored when person 3 steps up to fill the gap left by person 2. At time 10 there is a sudden increase in crowdedness as person 2 steps back and bumps into person 3. The graph in figure 1.2 is a way of representing the traveling **rarefaction** and **compression** waves shown in figure 1.1. Given a uniform acoustic medium, we could reconstruct figure 1.1 from figure 1.2 (though note the discussion in the next paragraph on sound energy dissipation). Graphs like the one shown in figure 1.2 are more typical in acoustic phonetics, because this is the type of view of a sound wave that is produced by a microphone – it shows amplitude fluctuations as they travel past a particular point in space.

An analogy for sound propagation

Figure 1.1 shows seven people (represented by numbers) standing in line to see a show. At time 2 the first person steps forward and leaves a gap in the line. So person two steps forward at time 3, leaving a gap between the second and third persons in the line. The gap travels back through the line until time 8, when everyone in the line has moved forward one step. At time 9 the first person in the line is shoved back into place in the line, bumping into person two (this is symbolized by an X). Naturally enough, person two moves out of person one's way at time 10, and bumps into person three. Just as the gap traveled back through the line, now the collision travels back through the line, until at time 15 everyone is back at their starting points.

We can translate the terms of the analogy to sound propagation. The people standing in line correspond to air molecules, the group of them corresponding to an acoustic medium. The excess gap between successive people is negative air pressure, or rarefaction, and collisions correspond to positive air pressure, or compression. Zero air pressure (which in sound propagation is the atmospheric pressure) is the normal, or preferred, distance between the people standing in line. The initial movement of person one corresponds to the movement of air particles adjacent to one of the tines of a tuning fork (for example) as the tine moves away from the particle. The movement of the first person at time 9 corresponds to the opposite movement of the tuning fork's tine.

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		1	1	1	1	1	1	1							
	1		2	2	2	2	2	2	X	1	1	1	1	1	1
*	2	2		3	3	3	3	3	3	X	2	2	2	2	2
	3	3	3		4	4	4	4	4	4	X	3	3	3	3
	4	4	4	4		5	5	5	5	5	5	X	4	4	4
	5	5	5	5	5		6	6	6	6	6	6	X	5	5
	6	6	6	6	6	6		7	7	7	7	7	7	X	6
	7	7	7	7	7	7	7								7

Figure 1.1 Wave motion in a line of seven people waiting to get into a show. Time is shown across the top of the graph running from earlier (time 1) to later (time 15) in arbitrary units.

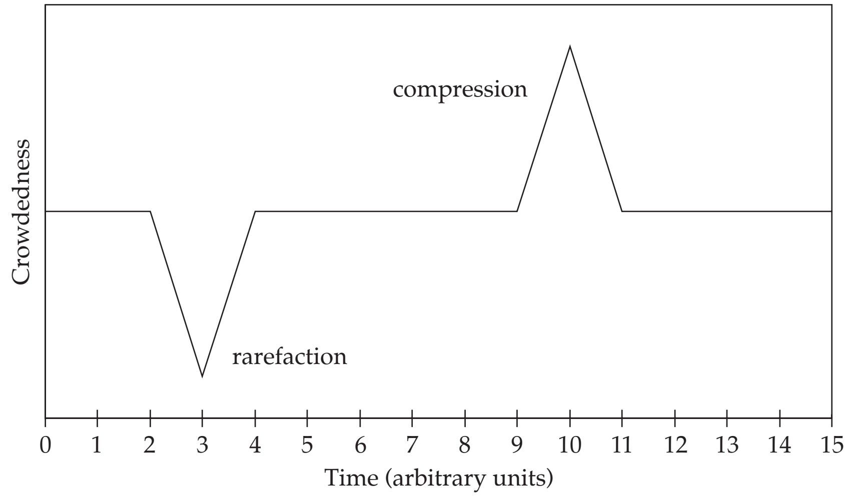

Figure 1.2 A pressure waveform of the wave motion shown in figure 1.1. Time is again shown on the horizontal axis. The vertical axis shows the distance between people.

Sound waves lose energy as they travel through air (or any other acoustic medium), because it takes energy to move the molecules. Perhaps you have noticed a similar phenomenon when you stand in a long line. If the first person steps forward, and then back, only a few people at the front of the line may be affected, because people further down the line have inertia; they will tolerate some change in pressure (distance between people) before they actually move in response to the change. Thus the disturbance at the front of the line may not have any effect on the people at the end of a long line. Also, people tend to fidget, so the difference between movement propagated down the line and inherent fidgeting (the signal-to-noise ratio) may be difficult to detect if the movement is small. The rate of sound dissipation in air is different from the dissipation of a movement in a line, because sound radiates in three dimensions from the sound source (in a sphere). This means that the number of air molecules being moved by the sound wave greatly increases as the wave radiates from the sound source. Thus the amount of energy available to move each molecule on the surface of the sphere decreases as the wave expands out from the sound source; consequently the amount of particle movement decreases as a function of the distance from the sound source (by a power of 3). That is why singers in heavy metal bands put the microphone right up to their lips. They would be drowned out by the general din otherwise. It is also why you should position the microphone close to the speaker's mouth when you record a sample of speech (although it is important to keep the microphone to the side of the speaker's lips, to avoid the blowing noises in [p]'s, etc.).

1.3 Types of Sounds

There are two types of sounds: periodic and aperiodic. **Periodic sounds** have a pattern that repeats at regular intervals. They come in two types: simple and complex.

1.3.1 Simple periodic waves

Simple periodic waves are also called **sine waves**: they result from simple harmonic motion, such as the swing of a pendulum. The only time we humans get close to producing simple periodic waves in speech is when we're very young. Children's vocal cord vibration comes close to being sinusoidal, and usually women's vocal cord vibration is more sinusoidal than men's. Despite the fact that simple periodic waves rarely occur in speech, they are important, because more complex sounds can be described as combinations of sine waves. In order to define a sine wave, one needs to know just three properties. These are illustrated in figures 1.3–1.4.

The first is **frequency**: the number of times the sinusoidal pattern repeats per unit time (on the horizontal axis). Each repetition of the pattern is called a **cycle**, and the duration of a cycle is its **period**. Frequency can be expressed as cycles per second, which, by convention, is called **hertz** (and abbreviated Hz). So to get the frequency of a sine wave in Hz (cycles per second), you divide one second by the period (the duration of one cycle). That is, frequency in Hz equals $1/T$, where T is the period in seconds. For example, the sine wave in figure 1.3 completes one

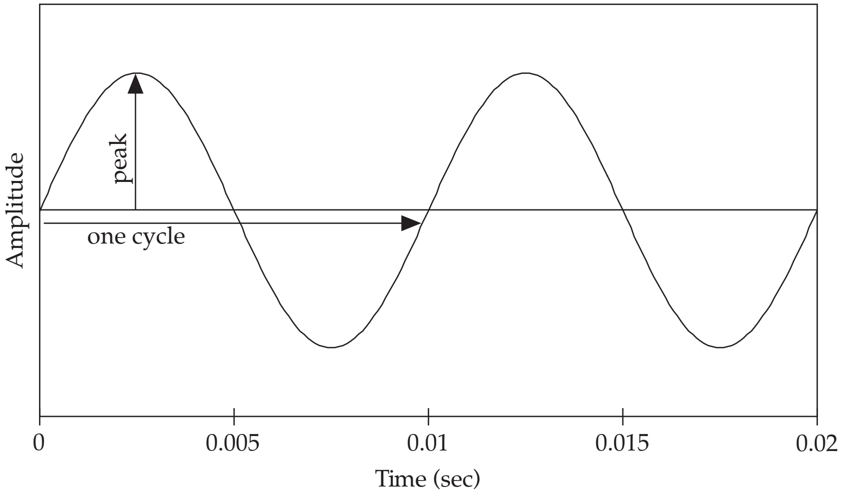

Figure 1.3 A 100 Hz sine wave with the duration of one cycle (the period) and the peak amplitude labeled.

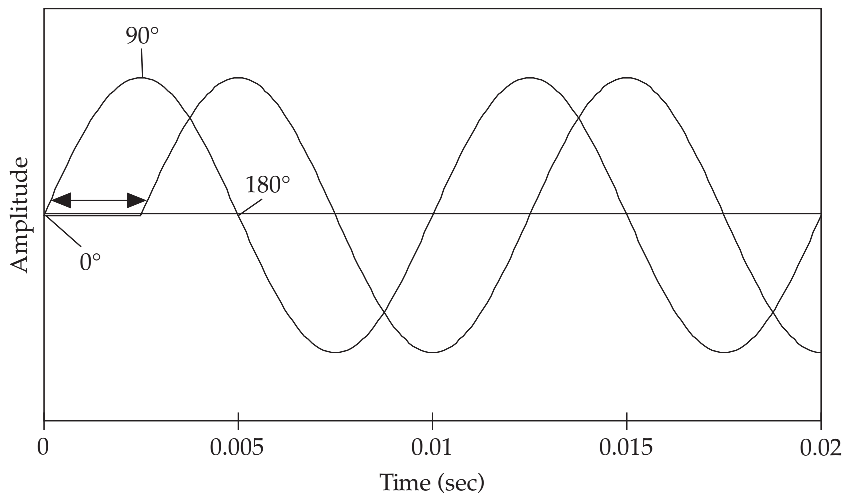

Figure 1.4 Two sine waves with identical frequency and amplitude, but 90° out of phase.

cycle in 0.01 seconds. The number of cycles this wave could complete in one second is 100 (that is, one second divided by the amount of time each cycle takes in seconds, or 1/0.01 = 100). So, this waveform has a frequency of 100 cycles per second (100 Hz).

The second property of a simple periodic wave is its **amplitude**: the peak deviation of a pressure fluctuation from normal, atmospheric pressure. In a sound pressure waveform the amplitude of the wave is represented on the vertical axis.

The third property of sine waves is their **phase**: the timing of the waveform relative to some reference point. You can draw a sine wave by taking amplitude values from a set of right triangles that fit inside a circle (see exercise 4 at the end of this chapter). One time around the circle equals one sine wave on the paper. Thus we can identify locations in a sine wave by degrees of rotation around a circle. This is illustrated in figure 1.4. Both sine waves shown in this figure start at 0° in the sinusoidal cycle. In both, the peak amplitude occurs at 90°, the downward-going (negative-going) zero-crossing at 180°, the negative peak at 270°, and the cycle ends at 360°. But these two sine waves with exactly the same amplitude and frequency may still differ in terms of their relative timing, or phase. In this case they are 90° out of phase.

1.3.2 Complex periodic waves

Complex periodic waves are like simple periodic waves in that they involve a repeating waveform pattern and thus have cycles. However, complex periodic waves are composed of at least two sine waves. Consider the wave shown in figure 1.5, for example. Like the simple sine waves shown in figures 1.3 and 1.4, this waveform completes one cycle in 0.01 seconds (i.e. 10 milliseconds). However, it has an