

HEARING SINGING

A Guide to Functional Listening
and Voice Perception

IAN HOWELL

Hearing Singing

Hearing Singing

A Guide to Functional Listening and Voice Perception

Ian Howell

ROWMAN & LITTLEFIELD
Lanham • Boulder • New York • London

Published by Rowman & Littlefield
An imprint of The Rowman & Littlefield Publishing Group, Inc.
4501 Forbes Boulevard, Suite 200, Lanham, Maryland 20706
www.rowman.com

86-90 Paul Street, London EC2A 4NE

Copyright © 2025 by The Rowman & Littlefield Publishing Group, Inc.

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the publisher, except by a reviewer who may quote passages in a review.

British Library Cataloguing in Publication Information Available

Library of Congress Cataloging-in-Publication Data

Names: Howell, Ian, author.

Title: Hearing singing : a guide to functional listening and voice perception / Ian Howell.

Description: Lanham : Rowman & Littlefield Publishers, 2025. | Includes bibliographical references and index. | Summary: "Ian Howell provides a fresh, actionable framework for the perception of the singing voice which will help guide singers toward efficient and expressive singing. The book dives deeply into the connections between voice acoustics, biomechanics, aerodynamics, functional listening, perception, and pedagogy"—Provided by publisher.

Identifiers: LCCN 2024051752 | ISBN 9798881804633 (cloth) | ISBN 9798881804657 (ebook)

Subjects: LCSH: Singing—Instruction and study. | Singing—Physiological aspects. | Singing—Acoustics. | Voice—Acoustics. | Auditory perception. | Musical pitch. | Tone color (Music) | Listening—Psychological aspects. | Spectral analysis (Phonetics)

Classification: LCC MT820 .H83 2025 | DDC 783/.043071—dc23/eng/20241101

LC record available at <https://lccn.loc.gov/2024051752>

™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48-1992.

To my teachers, my students, and my family

Contents

Acknowledgments

Preface

Before We Start

Who Is This Book For?

How to Read This Book

1 Introduction

The Purpose of This Book

The Problem

The Argument

The Structure of This Book

A Few Important Guardrails and Bits of Housekeeping

2 How We Got Here

The Burden of What Is Known About the Voice

Not All Singing Transmits Language

First Steps Toward Understanding the Timbre of a Voice

What We Choose to Notice

Conclusion

3 Refining Models for Understanding the Singing Voice

Current Pedagogic Models

How to Understand Phonation and Resonance in the Vocal Tract

How Does the Ear Work?

The Most Basic Model of the Ear

A Simple Model of the Ear

A More Complex Model of the Ear

Is Sound a Complex Wave or a Spectrum?

Introducing Pitch, Auditory Roughness, and Tone Color

Potential Conclusions and Applications

Discussion Questions

4 What Is Timbre?

What Differs? What Is the Same?

Time and Timbre

Potential Conclusions and Applications

Discussion Questions

5 Pitch

Pitch and Timbre

How Do We Perceive Pitch?

Where pitch is—and is not—in a spectrum

When Is a Fundamental a Fundamental?

Restating the Perceptual Qualities of Pitch

Unresolved Harmonics: The Noise in Pitched Sound

Issues with the Singer's Formant and Pitch

Beginning to Understand Pitch in Other Voices

Conclusions and Potential Applications

Discussion Questions

6 Time-and-Pressure Domain Considerations of Pitch and Timbre

Modeling Voice Production Without the Fourier Transform

The Timing of Timbre

What Is Formant Tuning?

The Pitch Exception of Overtone Singing

The Pitch Exception of Subharmonic Singing

The Timbre Exception of Vocal Fry

The Pitch Exception of Noisy Distortion

Conclusions and Potential Applications

Discussion Questions

7 Perceptual Qualities of Auditory Roughness and Pitch Resolution*

What Is Auditory Roughness?

Exploring Auditory Roughness Experientially

Visualizing Critical Bands of Hearing

The Equivalent Rectangular Bandwidth

The Intersection of Pitch Resolution and Buzziness in a Spectrum

The Two-Timbre Model and Beyond Potential Conclusions and Applications Discussion Questions

- 8** Tone Color and Brightness
 - What Is Brightness?
 - Tone Color: Conflicts between Speech and Musical Sounds
 - The History of Brightness
 - What If Formants Have Tone Colors?
 - Open Questions and Potential Applications
 - Discussion Questions
- 9** Absolute Spectral Tone Color
 - ABSOLUTE SPECTRAL TONE COLOR: HISTORY AND DEFINITION
 - Absolute Spectral Tone Color: Limitations
 - Exploring Absolute Spectral Tone Color Experientially
 - Vowels Share Common Tone Colors
 - Potential Conclusions and Applications
 - Discussion Questions
- 10** How Pitch, Auditory Roughness, and Tone Color Intersect: Theory and Application
 - The Timbre of a Formant as Pitch Changes: A Thought Experiment
 - The Intersection of Auditory Roughness, Pitch Resolution, and Tone Color In A Single

Formant
The Tone Color of “Flat” versus “Pointy”
Formants
Rules of Thumb
Some Examples of These Intersections
The Obvious True Fundamental
A Combined Perceptual Model
Perceptual Signifiers of Some Common
Resonance Strategies
The Tone Color of a Strong $1f_0$: Solutions for
the Treble Staff and Higher
The Tone Color of a Strong $3f_0$ or $2f_0$: Money
Notes across Genres
Hearing the Entire Spectrum
What Can Be Perceived?
Conclusions

11 How to Teach Singing
How a Singer Should Think
The Five Steps in All Singing
The Registration Triangle
A Few Practical Things to Keep in Mind
Is It Consistent? Is It Representative?
Different Parts of the Voice
What Does a Resonance Strategy Feel Like?
How the Entire System Works
Respiration

Phonation and Resonation
Acoustical Interactions
Summary
Good Rules of Thumb to Connect Perception to Function

- 12** Pedagogic Practices and Functional Listening Examples
Adjacent or Intermediate Functions and Exercises
Airflow Is Warmth / Warmth Is Airflow
Going on a Brightness Hunt
The Middle-Voice Slide
The Terrible Treble Ah
Same Shape / New Sound for Non-Trebles
Same Shape / New Sound for Trebles
Loud and Exciting Is Not Enough
There Are No Registers, Only Qualities
What Are the Perceptual Qualities of Vocal Registers?
What Are the Perceptual Qualities of Registration?
Thoughts on How Perceptual Qualities Point to Vocal Function
Some Basic Principles and Recurring Behaviors

Epilogue

[Glossary](#)

[Appendix A: How to Read Graphs Related to Voice Production](#)

[The Waveform](#)

[The Spectrum](#)

[The Spectrogram](#)

[Appendix B: List of Labs and Website URL](#)

[Bibliography](#)

[Index](#)

[About the Author](#)

Acknowledgments

This book is the result of my love for singing, my love for teaching others to sing, and my curiosity about the sound of the singing voice. I am grateful to have stumbled into the right set of people in the right order, each of whom bent my path just enough to point me toward connections that others seem not to have seen. This list is non-exhaustive, but it certainly includes Robert Cogan, Helen Greenwald, Katarina Markovic, Mattias Truniger, Donald G. Miller, Kenneth Bozeman, and Lynne Vardaman.

And I offer additional thanks to Daniel A. H. Mitton for his keen editorial eye.

Preface

BEFORE WE START

Welcome! I am glad you have found this book and that you are sitting down to ponder its contents. This book reflects who I am and how I understand the nature of singing and voice teaching. I hope you will feel welcomed into the way I think and that these ideas will find similar purchase in your mind.

I have long been fascinated by expressive, artistic singing, how one might deeply experience and understand that practice, and how one might effectively teach others to do it. I have lived all these lives as a performer, researcher, and teacher. I have won professional awards and received recognition for all three activities. But after working at this for decades, I am confronted with the reality that a deep understanding of the singing voice still lies just behind the veil. The human voice is understandable *in part* at different times in different ways. But the comprehensive nature of the voice is fundamentally unknowable and endlessly available for exploration.

In broad terms, the culture I find myself a part of does not always celebrate interdisciplinary work. Voice teachers and singers are certainly encouraged to incorporate conclusions from the scientific literature when they can be satisfactorily simplified. Toward this end there now exists a new education industry dedicated to the art of simplifying complex scientific information for voice teachers. However, a silent binary seems imposed: It is assumed that you must either be one who seeks to create knowledge or one who consumes that knowledge. Anyone interested in both may find themselves pushed away by voice teachers for being too science-curious and pushed away by scientists for being too practitioner-oriented.

However, in my experience encountering the voice pedagogy and vocology literature, some of the most interesting observations and

connections between the science and practice of singing have been made by those brave enough to attempt to straddle both worlds. This is the space occupied at different moments in recent times by William Vennard, Berton Coffin, Barbara Doscher, Cornelius Reid, Meribeth Bunch Dayme, Richard Miller, Donald Miller, Jeanie LoVetri, Jo Estill, and Kenneth Bozeman, just to name a few. These individuals were brave enough to absorb all they could from the resources available to them, ask the questions that only occur to those who sing and teach singing, and then turn around and suggest that they had a viable way to *think about singing* that was worth disseminating to others. It is perhaps bold to imagine the inherent value of your own point of view. But this is the only approach that changes minds. In my experience, it is the only approach that moves our field forward. In that spirit, I offer what follows.

WHO IS THIS BOOK FOR?

I can imagine several different kinds of readers who will approach this text. Some will bring a considerable understanding of the scientific literature on voice acoustics and perception. Some will come having been exposed to voice acoustics and perception in the context of a voice pedagogy or vocology class. Such an education may have attempted to cover what is *true*, what is *thought to be helpful*, or some combination of both. Some will come with no previous exposure to structured ways of understanding voice production, acoustics, and perception. And some may come with either indifference or active hostility to the idea that these scientific fields—or any kind of objective, quantitative measurement—can meaningfully inform the act of singing or voice teaching at all. It is difficult to simultaneously meet each of those very different populations where they are. As such, I will ask forgiveness from each of you in different ways and at different times for the way I have structured this book.

Ultimately, I will argue that we can learn how to *listen functionally*,¹ which is to say that we can associate specific aspects of the sound of a singer with specific physiological adjustments, and do so without the aid of a computer. This means we can use specific sounds as both indications of and prompts for specific functional outcomes. This is not a quick fix that distills well to bite-size takeaways. At least not if one wants to deeply understand why we hear singing the way we do.

Voice acoustics, physiology, aerodynamics, and perception are challenging to understand in part because the ideas these fields explore are inherently complex. Different people understand and teach them in different ways, which suggests that any one summary will leave out important information. Many refined and simplified

models emerge from these fields, but it is easy to mistake the ability to regurgitate the details of those models for a deep understanding of the underlying concepts. If we think of voice acoustics (in particular) in terms of computer measurements of singing, it is easy to imagine that one needs a computer to make use of it. If we lack a detailed understanding of important concepts from the study of sound perception, we may not notice important aspects of the sound a singer makes. This is true regardless of the detailed measurements we can take.

It is challenging to propose novel models for understanding the singing voice without engaging existing models. Some readers will have had thorough exposure to earlier models (and their associated tools), including the source-filter theory of voice production, the myoelastic-aerodynamic theory of phonation, the acoustic theory of vowel production, formant or resonance tuning strategies, acoustic registers, laryngeal registers, and, more broadly, the outputs of Fourier transform-derived spectra and spectrograms. Others will have never conceptualized the singing voice through such models or tools at all. If you are the former, please know that I will both challenge and seek to integrate those models in this book, and I will aim to bring nuance to the way you connect them to the act of singing from now on. If you are the latter, please know that any voice pedagogy or vocology textbook that addresses voice acoustics, perception, or production does so by using such models. And models are, by design, incomplete.

I do not expect you to join me on the academic ramparts as I lobby for changes in this literature. But if this is your first exposure to these ideas, I want you to be able to better understand the way in which texts that you read in the future about voice acoustics, physiology, aerodynamics, and perception selectively employ these models. When my arguments seem passionate, it is because I see much confusion that I believe could be alleviated by reframing

information within my communities. In my experience, many of the singers and voice teachers who wish to understand this material are stopped dead in their tracks by the details of the models themselves. If you do not see value in confronting those models yet, you will come to appreciate the importance of doing so as you continue to expand your library.

Those of you who seek to understand more about voice acoustics and perception may find the voice pedagogy and vocology literature confusing in two different ways. First, voice production is presented as either a time, flow, and pressure phenomenon or as a process of generating and filtering a spectrum of harmonics. The way in which these two ideas complement each other and reconcile is generally left unaddressed. A second challenge in this literature is the assumption that the output of such models maps well to qualitative aspects of the singing voice. It is well and good to note the volume of transglottal airflow per glottal cycle or look at a spectrogram and note the presence or absence of high-frequency energy. But I want to argue that we need to be able to discuss such phenomena in qualitative perceptual terms to be useful for singers and voice teachers. This means making actionable connections between our existing literature and models from the field of *psychoacoustics* (the study of sound perception). Otherwise, the best teacher would need the best computer, and there is clearly no historical support for that idea.

I am a strong believer that we notice the things we have labels for. Our brains thrive on predictive processes. This suggests that the act of learning how the sound of a singer might be broken down into a logical set of simultaneously present, qualitatively opposable percepts with dependably appropriate labels will change what we notice. In the simplest terms, this book proposes and provides experiential and scholarly support for a practical set of such labels

that allows one to listen functionally. This book aims to permanently change what you are able to hear.

HOW TO READ THIS BOOK

Depending on your background and the context in which you encounter it, I recommend one of three ways to read this book:

1. If you are interested in the deepest dive possible, and wish to ground ideas in academic frameworks, read this book from start to finish. Different parts will challenge you in different ways, but you will see the broad context of the culture that generates teaching materials for the voice pedagogy and vocology communities; you will be methodically taken through a granular understanding of perceptual phenomena relevant to the singing voice; and you will finish with a sense of how to apply those ideas in the voice studio.
2. If you are dipping your toes into the science of the singing voice, you may wish to start with [chapters 1](#) and 2 ("Introduction" and "How We Got Here") followed by the final two chapters ("How to Teach Singing" and "Pedagogic Practices and Functional Listening Examples"). Choosing this approach means you will have to accept certain claims I make without support for the time being, but this may equip you with the grounding you need to make sense of the middle chapters of the book.

3. If you are considering assigning portions of this book as readings in a voice pedagogy or vocology class, you may wish to use **chapter 3** as a broad overview of how to understand voice production, with **chapters 4** through 10 as special topic readings related to perception. **Chapters 11** and **12** will provide ample fodder for an applied teaching practicum.

No matter the approach you choose, please make use of the supporting resources I provide. In the appendixes you will find a primer on what the waveform, spectrum, and spectrogram images presented throughout represent; you will also find a glossary. Additional online resources are available at the following website: <https://www.embodyedmusiclab.com/hearing-singing> or <https://www.bit.ly/HearingSinging>. These include experiential lab assignments with video walk-throughs for you and your students.

It is my hope that you come away from this book with a clear sense of what qualities *can* be heard in a singing voice, that you have experienced these sounds, that you know how to support the assertions I make, and that you have a clear sense of how to link these sounds to the functional behavior of the singing voice. Good luck, and I will see you at the epilogue.

NOTE

1. Cornelius Reid, "Functional Vocal Training, *Journal of Orgonomy* 4, No. 2 (November 1970) and 5, No. 1, (May 1971).

1

Introduction

THE PURPOSE OF THIS BOOK

The sound of a singing voice is ephemeral yet powerful. Perfect harmonies can feel like physical objects in space. Dissonances rub, pulse, and yearn for resolution as though they are aware of their own discomfort. Singing can draw us into the present. It can take us back in time. How can this same act point us to the future or bathe us in nostalgia? How does one even begin to talk about the sound of a singing voice? *Can one talk about it?*

The purpose of this book is to attempt to address this question. Through an experiential and scholarly exploration of simple phenomena tied to sound perception, I will propose an actionable framework for *hearing singing*. This will take us through several subject areas which will be covered both academically and experientially. We will explore models for understanding the interactive physical and aerodynamic reality of voice production; we will explore how the ear both functions and plays a role in co-creating the sounds of our natural world; we will parse out simple and dependably identifiable aspects of timbre and perception; and ultimately we will find meaningful connections between those percepts and the underlying functional coordinations enabling the singer to produce them.

I have observed that the greatest compliment you can give a voice teacher is that they have “big ears.” The best teachers do not necessarily sing the best; they *hear* the best. This is a difficult skill to transmit, in part because words alone cannot convey the qualities we prize, nor can they ever completely summarize the numerous patterns that emerge across a student population. Effective training of voice teachers is a labor- and time-intensive process. As a small step toward addressing this challenge, I offer the book that follows.

Perception may be complex and difficult to pin down. However, there are simple ways of understanding the perception of sound (*psychoacoustics*) that can radically affect how you hear a voice. Acknowledging that the voice is complex, ephemeral, emotional, and culturally situated, I hope to convey to you that certain aspects of a singer's sound, understood with nuance and heard by trained ears, can convey technically actionable information to a voice teacher or coach.

THE PROBLEM

We have seen a profound expansion in our understanding of the physical world in the years since Manuel García II first viewed the oscillating vocal folds of a singer. This includes how the voice functions biomechanically, the nature of singing voice aerodynamics and acoustics, and our broader understanding of perception. The practice of integrating scientific information into voice pedagogy exploded in the nineteenth and twentieth centuries as content addressing biomechanics and voice acoustics became de rigueur in most serious voice pedagogy texts. However, the inclusion of these additional bodies of knowledge came with a corresponding burden. Authors were challenged to gain command of this specialized information and to effectively situate that information within a logical framework of questions and approaches to singing that remained culturally defined. The ways in which we came to understand these new bodies of information intersected with particular ways of singing.

Perception has long been considered prohibitively complex to tackle within the voice pedagogy and vocology literature. It has seemed too slippery to tie down, especially when contrasted with the quantitative measurements of anatomy and physiology, acoustics, and aerodynamics. One may find endless discussion regarding the existence (and computer measurements) of muscles, cartilages, bones, vibratory mechanisms, laryngeal and acoustic registers, and registration in a voice without a meaningful discussion of the perceptual nature of the sonic output.

THE ARGUMENT

This book argues for a fresh look at the perception of the singing voice, much in the same way that the nineteenth century ushered in our fascination with anatomy and physiology and the second half of the twentieth century deepened our collective understanding of the aerodynamics and acoustics of the singing voice. I hope this fresh look will help us to shift how we think about voice acoustics, biomechanics, aerodynamics, functional listening, *and* perception. I assert that these ways of understanding the singing voice are inextricably linked. I will explore how we can understand perception; introduce how we can teach it by tying what is known to how we experience sound; and explore how we can use these emergent connections to guide singers toward efficient and expressive singing.

THE STRUCTURE OF THIS BOOK

This book is organized into five sections plus one online resource. Many chapters end with a list of discussion questions that I hope will spark reflective learning.

1. I begin with a short introduction ([chapter 2](#)) to how the voice pedagogy and vocology literature has historically engaged information drawn from the scientific community. I will make an argument that we *should* care about singing voice perception, that we should confront the limitations of existing models that address it in the voice pedagogy and vocology literature, and that we can critically appraise the framework that currently houses these subjects in a voice pedagogy or vocology curriculum. This part of the book courts the philosophical, but it establishes an important framework for what follows.
2. I will introduce models to understand how the voice makes sound and how we hear that sound ([chapter 3](#)). This will include a discussion of popular models for teaching voice acoustics to voice teachers, and I will attempt to start to make connections between the details of those models and the physical reality of voice production.

3. I will next introduce and explore the intersection of three broad and actionable aspects of singing voice perception that are, in my experience, currently underexplored in the voice pedagogy and vocology communities: **pitch, auditory roughness, and tone color** ([chapters 4–10](#)). They are the signifiers of vocal function used in *functional listening*. More specifically, these three terms represent meaningful contrasts along the continua they reference: pitch/noise, pure/buzzy, and the vowel-like qualities between dark and bright tone colors.¹ Incredibly powerful aural cues lie at the intersection of these percepts as a singing voice migrates through changes in pitch, loudness, and timbre. This is especially important in the higher pitch and intensity ranges utilized in much singing (and typically avoided in speech). This part of the book will offer many lab assignments to explore these perceptual phenomena as you read about them.
4. The last two chapters will offer pedagogic principles, examples related to functional listening, and good heuristics for teaching voice ([chapters 11 and 12](#)). These chapters build upon the more scholarly presentation of models and concepts in the previous chapters, but that information is reformed into a narrative that ties

all the previous ideas together. In large part, these chapters lay out how I teach. As such, they situate functional listening within a practical pedagogical framework. If you are intimidated by the thought of exploring psychoacoustics in a technical manner, you may wish to start with these chapters.

5. Finally, I have included a glossary of difficult terms and two appendixes: a short primer that explains the various waveforms, spectra, and spectrograms that appear throughout this book, and a list of labs and that can be found at the indicated website. Throughout the book I will include references to these lab assignments. The website that houses them includes step-by step instructions and video walk-throughs.

A FEW IMPORTANT GUARDRAILS AND BITS OF HOUSEKEEPING

I have found that understanding the singing voice through a perceptual lens has profoundly impacted the way in which I think about and hear singing. This book is a thorough, slowly-sequenced exploration of that process and how I think about this subject. Please do not get bogged down in the prose when exploring ideas that could be made real with an experiential lab assignment. I will write under the assumption that you will seek out these sounds. None of this journey was abstract for me. It was always about the sounds.

The scope of this material *does not* extend to how a singer *fully* perceives their own singing body while singing, which is a fabulously complex phenomenon to consider. If we ultimately *hear with our brain*, that organ receives even more input as a singer than it does as a listener. My hope is to play a helpful part in demystifying this complex framework. The study of radiated sounds appears to be my best initial contribution.

Some of the information in this book is difficult because it asks for nuance around currently accepted models. For example, the idea that the vocal folds create harmonics, that “belt” *is* a strong second harmonic, or that a vowel’s dominant tone color is equally encoded across the spectrum. More on this later. I ask that you remain open to new ideas as they emerge and know that my objective is to present information in a helpful manner, regardless of whether it can be made simple enough for a novice.

This is not a comprehensive physics, perception, or acoustics book. And because I am interested in more than what is immediately practical in the voice studio, neither is it solely a voice pedagogy book. This text sits somewhere in between and is needed in part due

to translational issues that continue to arise between these communities. The ideas offered here enjoy a great deal of empirical and scientific support, have a long history of thought, and have stood the test of time in my own classroom and teaching studio. If my readers who are musicians are pushed to work harder to read this book than they are by other pedagogy books, and if my readers who are professional scientists raise an eyebrow when I sequence ideas for best consumption by the voice pedagogy and vocology communities, I will have hit the right balance.

I will adopt two notation conventions throughout this book. Pitch will be indicated with scientific notation. C4 is middle C. B3 is a half-step below and C5 is an octave above C4. I will also refer to the various aspects of the source-filter model after Titze et al. (2015).² Harmonics will be notated $1f_0$, $2f_0$, $3f_0$, etc. Resonances of the vocal tract will be notated f_{R1} , f_{R2} , f_{R3} , etc. I will also use this abbreviation to label slower and faster oscillations of air in the vocal tract. Formants will be notated F_1 , F_2 , F_3 , etc. What these terms point to in the context of different conceptual models will be explored throughout.

At many points throughout this book, I will reference the idea of a *pure tone*. In practical terms, this is analogous to a sine wave: a perfectly symmetrical oscillation of pressure that generates no harmonics. If you look at a spectrogram and see harmonics, they have the appearance of a sine tone. In the physical world, this is almost impossible to generate, as any inertia in the vibrating mass or resonating air will skew that pressure pattern. That skewing generates harmonics, even if imperceptibly. In the spirit of describing our experience of a sound over the measurement of a sound, I will tend to use the term *pure tone* to characterize very simple sounds. This sidesteps the question of whether the sound is completely

sinusoidal or not. When I use the term *sine wave*, it will be in the context of a mathematical process.

Finally, please keep in mind that I will suggest what is *possible* in terms of perception. I will point to what might be gained from revising our current conceptual and teaching models. It is impossible to receive this information in a vacuum, and I will likely push you to question the conclusions you have already successfully operationalized. This text challenges those models as much as it challenges the reader to experience the voice anew. Again, I encourage you to be patient and to ground the descriptions and explanations on the following pages in the sounds of real singing. Explore these ideas in your environment and allow yourself to be changed in the way that you hear a singing voice. This is going to be an ear-opening experience.

NOTES

1. Ian Howell, "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color" (DMA diss., New England Conservatory of Music, 2016); Ian Howell, "Necessary Roughness in the Voice Pedagogy Classroom: The Special Psychoacoustics of the Singing Voice," *VOICEPrints* (May/June 2017): 4–7.
2. Ingo R. Titze et al., "Toward a Consensus on Symbolic Notation of Harmonics, Resonances, and Formants in Vocalization," *The Journal of the Acoustical Society* 137, no. 5 (May 2015), 3005–7.

2

How We Got Here

Evidence-based voice pedagogy is a broad and ever-expanding field. Scientific studies slowly accumulate generally held knowledge about the voice. We also gather information from the lived experience of voice educators, and from the ongoing practice of exploring the limits of human voice production through singing.¹ Those working in the voice pedagogy industry have a sense that more is knowable—and that there are more ways of knowing—than any one person can fully grasp. Whether an idea is interesting, important, actionable, practical, or even applicable is frequently unclear, even after some passage of time. This means that much of the literature now under the umbrella of voice pedagogy rests on ideas that seem to be true and important, but that may or may not connect well to the real act of singing.

Vennard (1949) explicitly makes the argument for this approach in the first paragraph of his preface to *Singing: the Mechanism and the Technic*:

[This book] . . . is an attempt to compile under one cover objective findings from various reliable sources and to relate them to the art of singing. There are those teachers who feel that applying science to an art is quackery, but I believe

that our only safeguard against the charlatan is general knowledge of the most accurate information available. . . . The knowledge of literal facts is the only justifiable basis for the use of imagery and other indirect methods.²

This implies that the opposite of charlatanism is to ground teaching in scientific fact. However, this risks missing that (1) scientists create their conceptual models by reducing information in ways that are typically best suited to the ends of the scientific community, and (2) that scientific models are limited by the kinds of singing studied. These simplified models can, in turn, risk implying that the ultimate value lies in *knowing the fact* rather than *applying the principle*.

If greater exposure to information is an automatic good, we should be living in a golden age of voice teaching. Modern communication platforms now make it easy for anyone to share their take on voice pedagogy. One popular approach is to acknowledge the existence of a *science-derived concept*, further simplify and repackage it, and offer sound bites ready for the consumer. These layered simplifications are generally based on other, already simplified models, themselves used to further simplify and explain something even more complex. Such takeaways frequently reinforce the details of the model and do not meaningfully map to aspects of the underlying reality. This creates a bizarre paradigm where we may conceptualize the physical reality of voice production as though a scientific model is the starting point, rather than a summary of a complex, multistep analytical process. And this problematic process can still demand work on our part, creating a sense that we have learned something.

This is not a promising map to navigate in the twenty-first century. Well-curated information risks limited applicability across various ways of singing. More widely available information risks

being disconnected from the underlying phenomena summarized by scientific models.

THE BURDEN OF WHAT IS KNOWN ABOUT THE VOICE

It is an old idea to demonstrate mastery over a subject area for the purpose of developing one's own narrative as an authority. Harold Bloom (1997) terms this "the anxiety of influence," and it proves a useful framework beyond Bloom's field of literary criticism.³ Musicologists similarly understand compositions from the past as a weight that hangs on modern composers, demanding their acknowledgement and reaction.⁴ Voice pedagogy literature is frequently characterized by this tension. It is common to see new texts that incorporate existing and emerging research well before either the science is settled or the practical application is clear, because the existing body of material *must still be* acknowledged, appropriated, transformed, commented upon, and somehow reconciled within any new work. Without a credible connection to *what is known* more broadly, the new work may not survive the comparison. Off to the dustbin for the works of such charlatans, one might blithely suggest.

Here is an example: The publication of the first edition of *Gray's Anatomy* in 1858 speaks to the increasingly widespread dissemination of anatomical knowledge in the nineteenth century. Nineteenth-century pedagogues were forced to confront the emerging scientific fields of anatomy and physiology and to reconcile this information with their own views. This is not a bad thing, of course. However, Bloom might suggest that the nineteenth-century treatise authors who left anatomy and physiology unaddressed would have a harder time convincing their readers of their authority. For example, Giovanni Battista Lamperti uses specific anatomical language to indicate the appropriate location of sensations while singing: "As we said before, the resonance must be felt on top of the

head about where the *parietal bone joins the frontal bone (frontoparietal suture)*).⁵ This is a level of nuance and anatomical specificity—just say “the top of your head” if that is what you mean—that few of his students likely benefited from. This may seem like a historical curiosity until we ask ourselves how many modern voice pedagogy texts continue to foreground facts about the body as though exposure to information is an end unto itself?

Twentieth- (and now twenty-first) century voice pedagogy texts similarly confront the science of acoustics, made widely accessible by the research and publications surrounding the creation and distribution of the telephone. A modern voice pedagogy textbook chapter about voice acoustics would be incomplete without the spectra, spectrograms, formant plot graphs, or, later, the International Phonetic Alphabet and vowel quadrilateral generated by twentieth-century research into speech and sound transmission.

Assumptions about the utility of the acoustical models used to study the human voice have been passed down from generation to generation by authors who (perhaps unconsciously) incorporated older material to demonstrate their command of what was known. Transmitting the information represented by such models (for example, where various vowel formants are or what the contribution of the vocal folds is) does not mean that the teacher understands the implications any more than the student does. In fact, this lack of deep understanding may cause a voice pedagogy teacher to perpetuate their own experience as a former student: They teach the material because they feel they are *supposed* to understand it, whether they do or not.

Think about your own exposure to some basic concepts related to perception. For example, it is common to find statements in the voice pedagogy literature like: (1) Pitch is the cognitive experience of frequency; (2) timbre is everything besides pitch and loudness;

(3) different vowels have different formants; or (4) the ear carries out frequency filtering. Each of these statements is true at a model level. However, they do not necessarily tell you much about the nature of the phenomena they summarize or point to deeper implications.

Models for understanding scientific aspects of voice pedagogy have never been so widely shared, but the models themselves—which are generally simplified to characterize an *aspect* of singing within the context of a scientific study—may fail to connect to our sense of reality. Upon close inspection, they may hinder our ability to imagine the reality of a singing voice itself, promoting confusion and even backlash. Command of models, rather than the ability to make connections and generate effective outcomes, becomes currency. We may need to reevaluate Vennard's claim that experts are those who know facts and the rest are charlatans, when those who trade in facts they do not fully understand perhaps deserve that label as well.

NOT ALL SINGING TRANSMITS LANGUAGE

The human voice can transmit complex—even abstract—ideas through language, create purely musical or emotionally motivated sounds, or do both at the same time. Voice pedagogy borrows many of its models from the deep and fruitful fields of linguistics and language cognition (e.g., the International Phonetic Alphabet, vowel formant graphs, and spectrographic analysis). These models clearly bring powerful tools to bear on the study of the singing voice. However, these tools, developed to characterize spoken language, frequently fall short of describing the qualities prized in singing voices. The significantly wider pitch, intensity, and frequency spectrum ranges employed by singers may create physical phenomena that require novel analyses and descriptive models. These paralinguistic qualities are relevant to characterizing musical sounds, but they may have little to no impact on language cognition. They include distinctions along the continua of bright to dark, loud to quiet, high to low, buzzy to pure, tone to noise, and sound to silence. These are continua of qualitative oppositions that may transmit meaning within a single musical sound, gesture, and phrase irrespective of linguistic content.

Singing communities then have to generate a common, orally and aurally transmitted understanding of what is meant, felt, and heard when we discuss the *sound* of a singer, whether we discuss a clear or muddy vowel at a certain pitch, a specific registration choice, the difference between sufficient or insufficient glottal closure, or an overall color to define a genre. Within these communities, it is rarely a quantitative measurement that transfers the tradition. It is through sound, and many aspects of these sounds are quite separate from any linguistic meaning the singer may convey. By contrast, phonemes in speech tend to group subtly different sounds. These

internal differences, while meaningful, are generally not technically consequential to speech production. In singing, the physical adjustments associated with subtly different sounds may differentiate genres, emotional affects, or easefulness.

Anyone who sings in a style that requires deviations from a speech-shaped vocal tract to effectively resonate higher pitches (broadly called *vowel modification*) understands that singers leave the strict definitions of IPA symbols as pitch rises. Any singer who can sing above the treble staff similarly understands that high-sung pitches are qualitatively (timbrally) and experientially (what it feels like to produce) different from low pitches. Additionally, phonemes in speech are interdependent; it is more often the phonetic context—the sound of traveling from one phoneme to another—than the objective timbral character of a given phoneme that imbues a sound with linguistic meaning.⁶ However, singers frequently make sounds outside of linguistic context; what else is a sustained tone, melisma, vocalise, or wide-ranging riff but an exploration of the timbre and musical character of the singer's sound? It is perhaps more profitable to think of such changes in a singing voice as one would in a violin or piano, which are instruments with definite tonal characteristics based on pitch range, intensity, and emotional affect.

FIRST STEPS TOWARD UNDERSTANDING THE TIMBRE OF A VOICE

If one were to begin the study of the timbral qualities of a musical instrument with a clean slate, it is unlikely that the obvious first step would be to remove all sound below and above the analog telephone bandwidth (roughly 300 Hz to 3.4 kHz).⁷ However, with rare exception, that is exactly what our singing voice community does regardless of pitch, voice type, hormonal regime, performance environment, style, intensity, or affect. Our acoustical models for understanding the sound of a singing voice typically rest on either the frequency bandwidth of the first two spectral peaks (F_1 and F_2) or those two spectral peaks plus the classical singer's formant cluster. This turns out to correspond well to the frequency bandwidth of the analog telephone system. More on what these things are later.

This useful narrowing of frequency bandwidth allowed the telephone to preserve speech intelligibility while controlling costs and balancing resources. It also reflects that microphones and sound analysis technology at the turn of the twentieth century were unable to properly capture or study higher frequency energy.⁸ Somehow this narrow view of frequency information—born from practical considerations at the time—has turned into a narrative intimating that (1) the *extra* information below 300 Hz and above 3,400 Hz is either inaudible or irrelevant, and (2) that the information within that band may be lumped together. This has historically limited the scope of exploration.

But what may be perceived in the information that is routinely excluded from such analyses?⁹ What nuance may be brought to the information that is included in that narrow band? And how do these legacy decisions impact long-standing issues when such standards

do a better job of essentializing some bodies over others? As Tallon (2019) suggests:

The proliferation of AM (amplitude-modulated) radio stations in the early nineteen-twenties led to frequent signal interference, and by 1927 Congress decided to intervene by regulating the bandwidth allotted to each station. Both as a result of these limitations and advances in telephony research, most broadcasters and equipment manufacturers eventually limited their signals to a range between three hundred and three thousand four hundred hertz—a range known as “voiceband”—which was viewed as the bare minimum amount of frequency information needed to adequately transmit speech. Unfortunately, the researchers and regulators who were deciding on this range primarily took lower [speaking] voices into account when doing so.

Capping a signal at three thousand four hundred hertz didn’t significantly impact intelligibility for many men, but it certainly did so for most women, because it removed a significant portion of the sonic information critical for consonant identification.¹⁰

At best, this historical limitation falls well short of universally describing the singing voice. How could it when our voices can easily produce—and our ears can hear—sound outside this bandwidth? At worst, this approach points our ears and imaginations away from what may be pedagogically helpful ideas. If you would like to explore this experientially, complete Lab #1 (see appendix B for the website URL). Having reached this understanding of their origins, we may consider the benefit of discarding these limitations.

WHAT WE CHOOSE TO NOTICE

Perception is a curious phenomenon. It can be a nebulous, almost mystical process that is highly personal and dependent on education, context, and focus of attention. Listen to the same recording three times, and you can choose to hear something different on each pass. Listen to the same recording while paying attention to a passing butterfly or remembering the smell of your mother's macaroni and cheese, and your experience will change again. I want to suggest that it matters how you *choose* to perceive a voice. It matters how you *think*. Readers who work with singers have likely already decided to listen with greater specificity than they would as a novice. Any layperson who has acquired the taste for a given style of singing likely understands that their appreciation followed the decision to appreciate.

I would never argue that what follows is perceptually obligate. In general, people may listen to and enjoy a singing voice on their own terms. Much of the perception literature muddies this idea because the variable ways in which one might listen create the impression that perception is irredeemably varied. What I propose is a *way* to listen. I am confident that if you choose to listen in the manner I suggest, you will dependably notice the phenomena I point out. Furthermore, the way you hear issues in a singing voice will change, enlarging your toolbox in the teaching studio and refining your ability to listen functionally.

Ultimately, the aim of this material is to teach you what you can learn to notice through a simple set of heuristics. You will become aware of these qualities in singing, speech, musical instruments, and urban and natural sounds. These heuristics can come to characterize the fabric of your experience of timbre, which is as persistently available and as unseen as the air you breathe. In the following

chapters, I hope to make the argument for the utility of this model conceptually, pedagogically, and, most importantly, experientially.

CONCLUSION

I have laid out an argument for *why* we find so much science-derived information in our voice pedagogy and vocology culture. This information is there regardless of whether we deeply understand it. I have also set the stage to notice that the models we use to conceptualize the sound of a voice typically have their roots in speech research, and that singing and speaking are phenomenologically different. We now begin to explore these questions and these models.

NOTES

1. Kari Ragan, "Defining Evidence-Based Voice Pedagogy: A New Framework," *Journal of Singing* 72, no. 2 (2018): 157–60.
2. William Vennard, *Singing: The Mechanism and the Technic* (New York: Carl Fischer, 1967, iii; this excerpt also appears in the harder to find first edition from 1949).
3. Harold Bloom, *The Anxiety of Influence*, 2nd ed. (New York: Oxford University Press, 1997).
4. Joseph N. Straus, "The 'Anxiety of Influence' in Twentieth-Century Music," *The Journal of Musicology* 9, no. 4 (1991): 430–47.
5. Giovanni Battista Lamperti, *The Technics of Bel Canto* (New York: G. Schirmer, 1905), 15. Emphasis added.
6. See Mark H. Ashcraft, *Cognition* (Upper Saddle River, NJ: Pearson Prentice Hall, 2006), 382–83, for a discussion of coarticulation. See also Terrance M. Nearey, "Static, dynamic, and relational properties in vowel perception," *Journal of the Acoustical Society of America* 85/5 (May 1989): 2088–113, for a discussion of conflicts between the context effect and inherent quality in speech research.
7. Brian Monson, Eric J Hunter, Andrew J Lotto, and Brad H Story, "The Perceptual Significance of High-Frequency Energy in the Human Voice," *Frontiers in Psychology* 16, no. 5 (2014), 2.
8. Monson et al., "The Perceptual Significance," 2.
9. Monson et al., "The Perceptual Significance," 1–10; Ingo R. Titze and Sung Min Jin, "Is There Evidence of a Second Singer's Formant?" *Journal of Singing* 59, no. 4 (2003): 329–31; S. O. Ternström, "Hi-Fi Voice: Observations on the Distribution of Energy in the Singing Voice Spectrum above 5 kHz," *The Journal of the Acoustical Society of America* 123, no. 5 (2008): 3171–76.
10. Tina Tallon, "A Century of 'Shrill': How Bias in Technology Has Hurt Women's Voices," *New Yorker*, September 3, 2019, <https://www.newyorker.com/culture/cultural-comment/a-century-of-shrill-how-bias-in-technology-has-hurt-womens-voices>.

3

Refining Models for Understanding the Singing Voice

The singing voice is too complex to grasp all its details all at once. Instead, we use models to teach important, generalizable concepts. Because it is the nature of all models to exclude information, the art of creating a good model lies in knowing which information to leave out. By definition, models are *missing* information. A single model that has been designed to characterize an aspect of the voice is only ever an incomplete way to think about singing.

At the heart of this chapter is a very basic set of questions: Do models based on the source-filter theory—where the vocal folds generate a spectrum of harmonics filtered by the vocal tract—capture the physical nature of voice production, or do they pull voice teachers away from a deeper understanding of vocal function? How does the ear transduce the information in sound waves, and how does that process affect what we hear? Does the output of a spectrum or spectrogram adequately represent the perceptual qualities of a voice? And finally, are there other, equally viable models we could deploy in place of the source-filter model?

In teaching voice acoustics—and I will start with voice acoustics because it generates so many conceptual models—we lean heavily upon spectra and spectrograms and the language of the source-filter theory of voice production. These images show the results of a *Fourier transform*. The Fourier transform is a mathematical process that decomposes a

complex pattern (an audio recording, for example) into simpler waveform components with specific frequencies, amplitudes, and phases. By combining the frequency, amplitude, and phase of each simple waveform component, we can redraw the complex waveform. If the audio recording is periodic—which means it likely has a pitch—the Fourier transform will return a spectrum of harmonics of varying intensities. The frequency and amplitude components are usually displayed in this spectrum, while the phase information is hidden. The resulting harmonics (prominent simple components generated by the Fourier transform) commonly form the basic visual language that we use to conceptualize resonances, formants, vowels, and the other elements that form the acoustic signature of a voice. Despite conceptual similarities, the ear processes complex, periodic tones differently than a Fourier transform. In this chapter, I will begin to explore the ways in which these two processes produce different results.

I believe that voice teaching is, at its core, a heuristic pursuit. There is no one piece of information that could be found in a voice pedagogy or vocology book that is *automatically* helpful. The razor of practical application is ready to dismiss ideas that seem to bear no immediate practical benefit to the training of a voice student. And this is likely as it should be: Hard truths bend to the practical needs of real human singers who are seeking solutions amid their complex emotional and physical experiences of life. But it means that our practical teaching models are typically only as accurate as they need to be in order to get the job done given what our target audience already understands. Our models are sufficient and helpful rather than accurate or complete. Again, most would argue this is as it should be. However, this does mean that our field is structurally vulnerable to centering problematically incomplete models.

The models we use to explain singing tend to be organized by the systems into which we subdivide the singing body. We tend to discuss the singing body in terms of these discrete subsystems: respiration, phonation, resonance, and articulation. This popular organizational structure leaves out two important subsystems: radiation (the ways in which the sound of a voice changes because of the room) and perception (the act of hearing). Additionally, it frequently leaves out the role of the

brain in instigating the singing in the first place, but that is largely beyond the scope of this book.

I wonder how our corporate understanding of the singing voice would change if these pedagogical models accommodated what is already known from the study of sound perception (*psychoacoustics*). Psychoacoustics offers important tools to explore voice production. Rather than simply observing the objectively measurable aspects of a singer's audio signal, we can begin to ask more actionable questions such as:

- What does that aspect of the singer's spectrum sound like?
- What does it contribute to the whole?
- Does that part of the spectrum represent pressure and flow patterns that play a role in sustaining phonation?
- Does that part of the spectrum stimulate my ear in a way that the act of perception dependably introduces qualitative elements of timbre?
- Can I train my ear to listen for those qualities and then link them to successful or unsuccessful attempts at a technical or functional goal?
- Can I hear everything I see on a spectrum or spectrogram?
- Can I use this information to revise the models I currently use to better conceptualize, or better teach others to understand the behavior of the voice?

As I wrote in the last chapter, the field of voice pedagogy has already assimilated ideas from anatomy, physiology, linguistics, and acoustics. We are currently attempting to do the same with motor learning theory, learning theory, and neuroscience. In cases where such synthesis has proven useful, the interdisciplinary mingling can provide additional support and explanations for what works well, can generate new teaching frameworks, and in some cases can even reform what have been questionable assumptions. Open any modern voice pedagogy book to see how these topics delineate chapters. Whether the information therein successfully connects to the needs of the singers and teachers who read it is another question. What follows is meant to enliven what works well in our current models, while seeking to account for some of the unsatisfying

exceptions and gaps in logic that our existing models currently accommodate.

CURRENT PEDAGOGIC MODELS

Most science-based pedagogic models for thinking about the singing voice will center on either the neurological, physiological, and aerodynamic means of voice production or the acoustical measures of its output. The former is certainly an important scaffold on which voicing hangs. It allows one to consider how specific muscles contract to move the various parts of the singing body, and especially how the vocal folds behave in and act on the flow of air. Acoustical measures, on the other hand, perhaps require even more finesse to understand.

Most acoustical measures are derived from the mathematical decomposition of the singer's radiated pressure wave as seen in a spectrum or spectrogram. A spectrum is used to average the rates of pressure change within that radiated wave. A spectrogram is a time series of those spectra. In digital computers, this decomposition is typically accomplished by means of either a discrete or fast Fourier transform (DFT or FFT), the complex mathematical processes under the hood in any spectrograph. Most pedagogically actionable conclusions rest on the relationships between the visible structures in the resulting spectral images. It is worth asking whether such a model shows what is *real* or whether it shows *aspects* of a singer's sound that could be understood in other ways. To put it directly, what *does* a spectrum or spectrogram show? Does what it displays physically exist in the air? In the listener's ear? In the listener's percept of the sound? If perception is a response to the sound itself, understanding the acoustical process that generates that sound is the place to start.

HOW TO UNDERSTAND PHONATION AND RESONANCE IN THE VOCAL TRACT

If we are to link what we see in a spectrum to the reality of singing, let us begin with a simple overview of the physical means of voice production and its acoustic output. It may be helpful for the reader to consider that voicing is a multistep process where each step transforms an actual, physical phenomenon. The pressurizing of tracheal air in many ways represents the *potential* to voice; however, one cannot layer complex aesthetic qualities onto it. The train of air puffs that escape the rapidly opening and contacting vocal folds during phonation are perhaps most simply modeled by a line graph charting transglottal airflow against time.¹

The pressure dynamics below, between, and above the vocal folds are also important to understanding the nature of self-sustaining vocal fold oscillation and vocal tract resonance. In normal phonation, air flows from below to above the glottis when the vocal folds are open, and this airflow abruptly drops when they contact. In cases where the vocal folds completely close, this airflow drops to zero.

This flow pattern meets a mass of air in the vocal tract, which has inertia. This means that the vocal tract air mass typically responds slightly sluggishly to the “new” air, which causes a region of supraglottal high pressure to build up. When the glottis closes—or reaches maximum contacting, depending on the phonation quality—several important things happen:

1. The significant drop (or cessation) in airflow conspires with the higher-pressure supraglottal air mass already moving away from the vocal folds to generate an area of low supraglottal pressure. It is not a true vacuum, but this air rapidly drops well below atmospheric pressure. It may be confusing to make this connection; why would shutting off airflow make the

pressure drop? Consider that pressure here is a measure of the amount of air in a space. It is density. If the air above the glottis is moving away and no new air is introduced to replace it, the density (and therefore pressure) drops.

2. When the vocal folds are in contact, the vocal tract becomes a more effective resonator. In simple terms, if you understand resonance as the bouncing of pressure waves back and forth between the base of the pharynx and the opening of the mouth (or nose), that process is more efficient if you do *not* allow that pressure wave to enter into and be absorbed by the lungs. Since the contacting vocal folds reduce the rate of airflow faster than they ramp up the rate of airflow as they are opening, the resulting drop in supraglottal pressure happens faster than the preceding rise. It is this pressure change (from high to low) that propagates through the vocal tract, setting that air mass into motion. I will refer to this pressure drop as an impulse, and it is acoustically like the sudden change in pressure that generates the sound of a handclap or the sound of a popped balloon. As the folds are in contact at this moment, the response of the vocal tract remains stronger for longer than if the vocal folds had been open.

For readers accustomed to conceptualizing the contribution of the vocal folds to phonation in terms of discrete harmonics—a tenet of the source-filter model—it may be difficult to reorganize this basic concept. However, the slower rise and faster fall of supraglottal pressure as the folds open

and then contact represents continuous and accelerating changes in pressure. This means that somewhere along that accelerating pressure curve, a wide range of frequencies (rates of pressure change) exist. Hence the vocal folds work with the vocal tract to generate a broadband (all frequencies) impulse in the supraglottal air mass. If this is counterintuitive, record the sound of a handclap with a closely placed microphone and view the result in a spectrogram. That impulse similarly shows energy from very low to very high frequencies.

Like a length of PVC pipe responding to a sudden slap at one end or a reverberant room responding to the pop of a balloon—acoustically if not aerodynamically—the vocal tract reshapes this broadband impulse by the means of its own standing resonant modes of vibration.² Another way to say this is that the air mass in the vocal tract is best able to move at certain speeds, based on its shape. The frequency ranges (rates of change in pressure) from the source impulse that coincide with those vocal tract resonances oscillate in pressure over time with greater strength than those that misalign. Frequency regions that misalign are quickly attenuated. What was a simple, repeating slow rise and fast fall in pressure facilitated by the modulation of transglottal air by the vocal folds becomes a substantially more complex pattern with multiple rises and falls in pressure of varying amplitudes. We might even say that the tube and the reverberant room change the *timbre* of each broadband impulse by allowing certain patterns and speeds of pressure change to ring for longer (continue to oscillate) than others.

Here is another way to think about this: The vocal tract reflects the energy of that broadband source impulse at various changes in diameter along its length. These reflections all interfere with one another, both constructively and destructively. Some frequency regions (rates of change) are allowed to continue oscillating, while others are quickly attenuated. The total set of vocal tract resonances emerge as the sum of this complex cascade of interacting reflections within the vocal tract.³ Hence there is no single location in the vocal tract that *is* the source of a given resonance. The frequency ranges that survive this total wave

interference—which we see as formants in a spectrum—emerge from the initial pressure wave as it decays.

If we sit with this idea for a moment, it may occur to you that the *filter* portion of the *source-filter model* may do more than filter the source. Does a wall filter the sound of a handclap when it produces an echo, or does the wall make a copy of that pressure wave and change it consistent with the acoustical properties of the wall? Stand in a corner facing into a room and you will notice that the reflections of your voice off the walls provide a strong boost in both warmth and intensity. Did these walls not add their own reflections to your direct sound, generating a new sound augmented by the additional sound sources? And if the vocal tract produces a cascade of reflections to the source impulse, are those not new pressure waves? Put another way, is the vocal tract itself a sound source, albeit one that requires input to function? Heller (2013) goes so far as to suggest:

The source-filter idea is technically wrong, as long as you think of a filter as something that removes things, but it persists legitimately as a *model*. It seems a shame, however, to deprive singers and speakers of the notion of resonance, and instead leave them with the flat-footed impression that they are merely filtering whatever their vocal folds put out.⁴

Indeed, we may think of the vocal tract as a grouping of individual resonances, and one can dependably effect change in just a portion of the radiated spectrum through specific articulatory manipulations.⁵ Brad Story (2019) suggests that the entire vocal tract does appear to act on the entire source broadband impulse, although he frames it in spectral terms.⁶

When this pattern of impulse and resonance decay repeats periodically, as is found in a singing voice, pitch is layered onto timbre (more on this in [chapter 5](#)). Before the second impulse interferes with the resonant decay of the first, and the third similarly of the second, the source impulse theoretically contains the potential for much lower frequency (slower oscillations) information than most singers could sing as a pitch. It is the

repetition of the source impulse that filters out information that unfolds slower than the duration of the period, disallowing lower frequency information from emerging (see [chapter 6](#) for a more thorough exploration of this concept).

It seems credible to think of the transformation of the source signal by the vocal tract as the transformation of a broadband impulse generated by asymmetrical oscillations of airflow into a complex pressure wave that takes place per period of voicing. Note that viewing resonance as a function of time *within a glottal cycle* will ultimately sidestep the idea that all the harmonics shown in a spectrum persistently exist at all timescales.

The reshaped (complex) pressure wave radiated from the singer is further transformed by the acoustical properties of the room (and the signal chain if the singing is amplified),⁷ the listener's body, their pinnae, and their ear canals.⁸ Once the energy of this complex wave is transferred from the last line of air molecules in the ear canal to the eardrum, the physical limitations of the middle and inner ear come into play.

What we have just explored here is different from common acoustical models in voice pedagogy, which suggest the interaction of voice source harmonics and vocal tract resonances.⁹ To be clear, no physicist or acoustician would likely suggest that this harmonic model is strictly accurate. But my long participation in the science-informed voice pedagogy community tells me that this is how many singers and many voice teachers think. Admittedly, this model is heuristic and helpful; however, it risks profoundly distorting our sense of how the system works.

Even still, every step of this process *can* be understood through the lens of a Fourier transform, which *appears* to imply the opposite: that the vocal folds create a periodic source sound independent of the resonator, that this source sound can be understood as a buzzing sawtooth wave, and that that physical phenomenon is equivalent to a spectrum of distinct harmonics.¹⁰ This widely accepted and broadly taught model appears to imply that the signal fed into the vocal tract *is* a set of separate harmonics that interact independently with similarly independent vocal tract resonances.

If we believe that the complex, physical wave generated by a voice is a collection of harmonic pure tones—which is what a spectrum at least initially appears to show—then to operate within this model we must also mismodel the ear as a system capable of receiving that input. As I will explore below, the ear cannot do this. Ultimately the physical wave generated by the voice is a complex pattern of pressure changes that move through space over time, regardless of how one might mathematically summarize it. Since this is so, it follows that every implementation of this harmonic model in the voice pedagogy literature is at least partially incomplete. Correspondingly, the way that voice acoustics is taught in the voice pedagogy classroom, the way that modern voice educators disseminate this information online, and the way that well-intentioned voice teachers try to implement these concepts in the studio all may suffer from this framing.

By contrast, the source-filter model of speech production rests on the notion that this is a reasonable way to conceptualize the system.¹¹ And it is a remarkably useful model. It is true that at any point in the physical process of voicing explained above, one could impose a Fourier transform on the signal and consider the pressure wave according to its spectral components. However, this does not mean that a spectral analysis captures either the physical nature of the phenomenon at a given moment, or that all it reveals are perceptually relevant structures.

I posit that this is a good example of an overlooked distinction: Voice pedagogy texts present a reasonable way to *model* discrete parts of the voice, but these well-meant efforts risk centering an incomplete description of how the voice *works*. We can think of the sound of the oscillating vocal folds sans vocal tract; with a cadaver larynx and an air compressor, we can actually record that sound, which does sound like a buzzing sawtooth wave—a sound with a complete spectrum of harmonics that drop in intensity the higher in the spectrum one looks.¹² But that sidesteps the fact that each individual impulse of energy arising from the closing cadaver folds excites the unbound air above the excised larynx per glottal cycle. There is no vocal tract in this experiment. And many different periodically repeated patterns will generate a spectrum of

harmonics when passed through a Fourier transform. In other words, this experiment does not demonstrate that the vocal folds generate harmonics when connected to a vocal tract. It demonstrates that they produce a periodic pattern that can be analyzed in terms of harmonics, even without a vocal tract.

In vivo, the sound generated by an excised larynx never actually exists as a repeated phenomenon. It is always coupled with the vocal tract. We can think of the effect of the vocal tract in terms of what would have to interface with each individual source harmonic. But that would require that we conform the physical behavior of the vocal tract to accommodate this specific model of the input itself. Finally, we *could* assume that the images that commonly configured spectra and spectrograms show—with discrete harmonics and formant peaks—accurately represent the way in which we perceive a voice. But this ignores the transformative power of the ear.

HOW DOES THE EAR WORK?

Exploring the function of the ear is helpful for at least two reasons: (1) I make claims throughout this book that the ear mediates the stimulus that it receives by introducing perceptual qualities unrealized in the raw pressure wave in the air. I suggest that we can understand the timbre of a singing voice in these terms. (2) An important argument in this book is that the study of voice acoustics breeds confusion in part because of the models we use to teach it.

Like any other biological system, one may understand hearing through a series of models that become more accurate as they become more complex. The following text lays out three models of the ear that take us from a quite basic understanding to a suitably complex one. Keeping in mind the importance of synthesis, I will discuss the relevance of including the complex nature of the acoustic output of the voice in our pedagogic model.

The Most Basic Model of the Ear

In basic terms, the ear receives pressure waves from the environment and conveys the information contained in them to our brain. A young human with pristine hearing can hear sounds in the frequency range between 20 Hz and 20 kHz. Some might dub this a “black box” model because we have an input and an output without a clear sense of what takes us from one to the other.

A Simple Model of the Ear

Most simple models divide the ear’s anatomy into three parts: the *outer ear*, the *middle ear*, and the *inner ear* (see [figure 3.1](#)). The outer ear includes what may be understood as the external, cotton-swab-accessible structures: the *pinna* (the external structures) and the *ear canal* (also known as the *auditory canal*). This canal terminates at the outer wall of the *eardrum* or *tympanic membrane*. These external structures gather

vibrations from the surrounding environment (typically through air), which in turn set the tympanic membrane in motion.

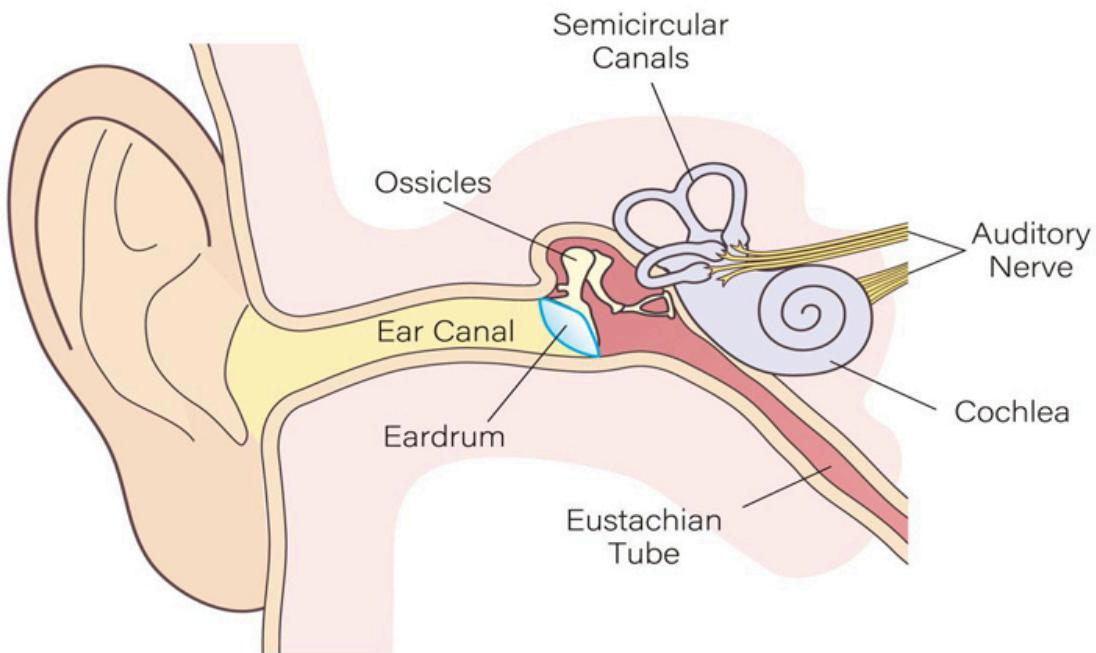


Figure 3.1 Gross anatomy of the ear. Source: yumiimage (adobe.stock.com)

The middle ear consists of three minuscule bones called the *ossicles* (the *malleus*, *incus*, and *stapes*) and their various connective tissues. These bones transfer vibrations from the inner surface of the tympanic membrane to the structures of the inner ear. What had been pressure changes in air become the mechanical motion of the ossicles.

The inner ear is primarily a fluid-filled structure called the *cochlea*, which is part of the same bony labyrinth structure that contains the vestibular organs that are responsible for our sense of balance. The cochlea itself looks somewhat like a snail shell. Vibrations from the bones of the middle ear excite small structures within the cochlea that filter sounds into the frequency spectrum we are familiar with from spectra or

spectrograms. The cochlea converts these stimuli into electrical impulses which are sent to the brain.

A More Complex Model of the Ear

A more detailed model could consider the shape of one's shoulders, head, and pinna as factors in gathering pressure waves from the environment. The auditory canal has its own resonance centered approximately between 2,700 and 3,300 Hz,¹³ and pressure waves are reshaped as they pass through on their way to the eardrum. The motion of the tympanic membrane in response to incoming changes in air pressure sets the ossicles into a pattern of mechanical vibration. What had been acoustic energy—vibrations of air molecules—transforms into the mechanical vibration of these tiny bones that are free to move along one rotational axis each. Their physical characteristics amplify energy in the 1,000 Hz to 1,500 Hz range.¹⁴

The combined effect of the resonances of the outer and middle ear (the auditory canal and the ossicles), significantly increases the sensitivity of the human ear to a wide range of frequencies between approximately 1,000 Hz and 3,500 Hz. This is the range of many of the vocal tract resonances favored in speech. We see the effect of this in the equal-loudness contours (see [figure 3.2](#)) with notable dips (increased sensitivity) at 1,000 Hz and 3,000 Hz and a general increased sensitivity between approximately 300 Hz and 5,000 Hz. In this image, greater sensitivity is indicated by a lower vertical axis value. The ear is dramatically less sensitive at very low and very high frequencies (a higher vertical axis value).

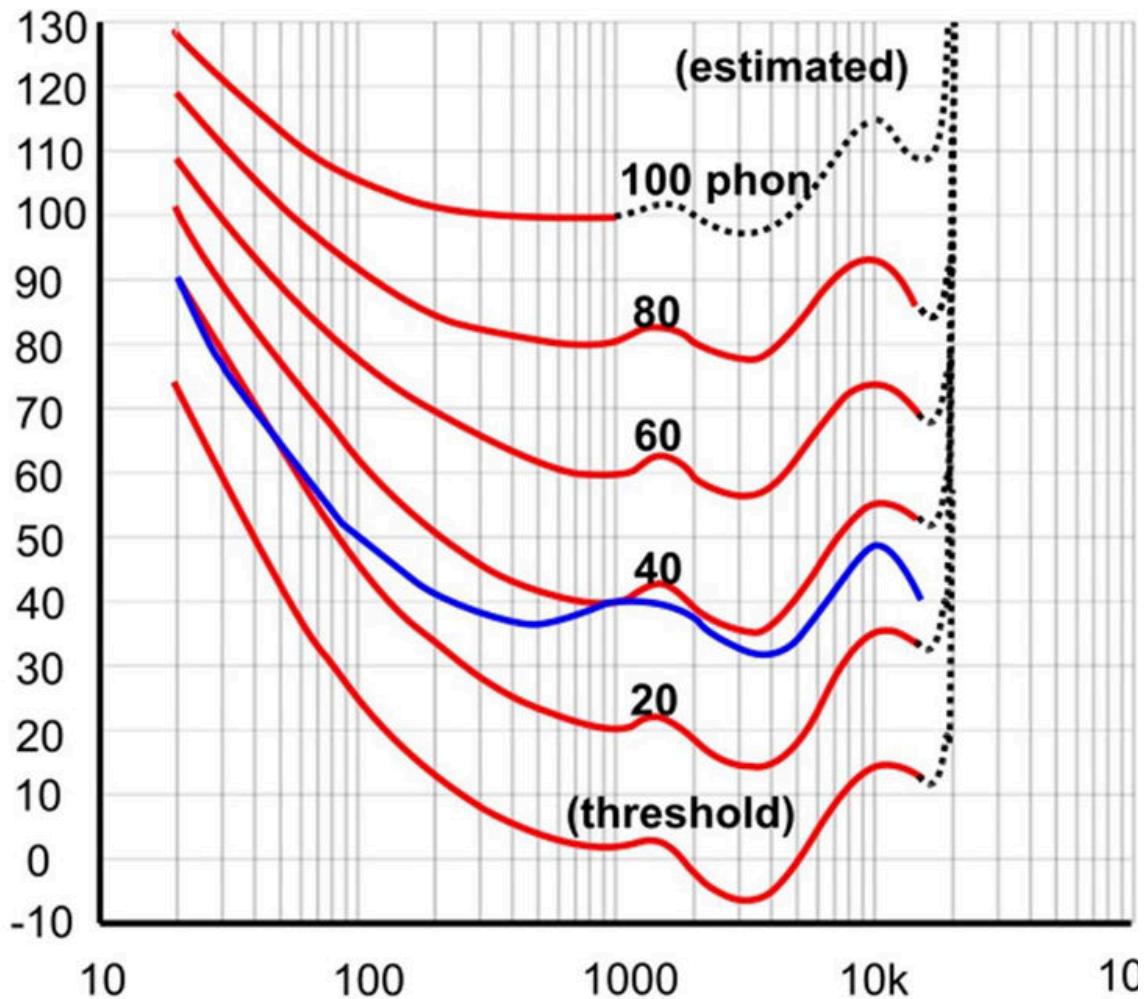


Figure 3.2 The Equal Loudness Contours based on the ISO 226: 2003 standard. Frequency (Hz) is shown on the horizontal axis and sound pressure level (dB) on the vertical axis. Source: https://commons.wikimedia.org/wiki/File:Equal_loudness_no_caption.svg

These sensitivity differences between low and high frequencies grow smaller the more intense the stimulus is (contours that start at a higher vertical value). They grow larger as intensity drops. In what follows, I will argue that specific, qualitatively opposable aspects of the timbre of the radiated sound of a voice correlate well with specific functional adjustments of the singing body. Our sensitivity to these *signifiers of functional listening* will vary with distance to the singer and intensity of their singing. This means that the sound of their voice will differ in a small

studio compared to a concert hall and that it may be easier to hear these important qualities when physically closer to a singer. This also means that an amplified singer may offer one a more direct sense of the underlying vocal function as their signal is captured prior to much propagation in a space.

The mechanical vibrations of the ossicles terminate at the footplate of the final bone, the *stapes*. This footplate acts like a piston that pushes into and pulls out of the cochlea, setting the cochlear fluid into motion. The area of the tympanic membrane is greater than that of the footplate of the stapes. This means the ossicles not only transfer the mechanical vibration of the tympanic membrane to the inner ear but also step up the amplitude of the pressure wave introduced into the fluid-filled cochlea. This overcomes the higher resistance (impedance) of the cochlear fluid compared to the less dense air on either side of the tympanic membrane.¹⁵

The cochlea (see [figure 3.3](#)) contains two main fluid-filled tubes—one situated adjacent to the other—that spiral from the base to the apex of its structure. The ascending tube (base to apex) receives the vibration of the stapes and is called the *scala vestibuli*. The descending tube (apex to base) is called the *scala tympani*. Between these two tubes is a third, called the *cochlear duct*. The boundary between the scala vestibuli and the cochlear duct is called the *vestibular* (or *Reisner's*) *membrane*, and it allows pressure waves from the scala vestibuli to enter the cochlear duct. The boundary between the cochlear duct and scala tympani is a variably flexible structure called the *basilar membrane*. The basilar membrane plays an active role in frequency-filtering pressure waves as they propagate through the cochlear duct.

Anatomy of the Cochlea

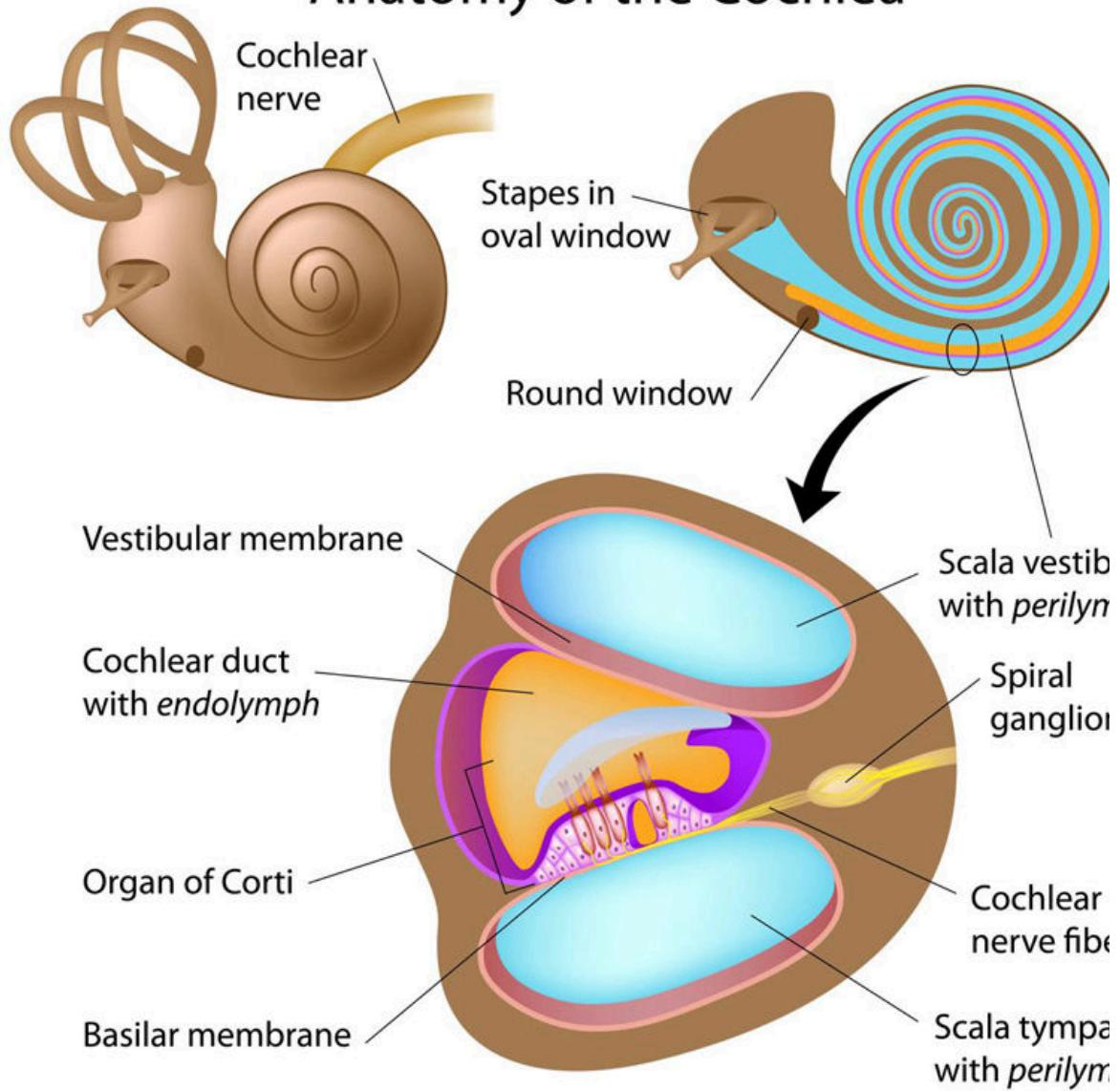


Figure 3.3 Detail of the cochlear anatomy. Source: Alila Medical Media (stock.adobe.com)

Once inside the cochlear duct, those pressure waves pass through the basilar membrane to the descending tube (scala tympani), which returns the energy to the base of the cochlea. This tube terminates in a flexible tissue relief valve called the *round window*, which allows for the displacement of the fluid in the scala vestibuli and scala tympani by the

stapes. This fluid is called *perilymph*. The fluid in the cochlear duct is called *endolymph*.

The response of the basilar membrane to pressure changes is organized tonotopically, which literally means “frequency place.” Near its base (where the footplate of the stapes introduces vibration into the scala vestibuli), the basilar membrane is sensitive to high-frequency sounds. Toward the apex, it is sensitive to low frequency sounds (see [figure 3.4](#)).¹⁶ In broader conceptual terms, we have once again arrived at a model that appears to align with what we see on a spectrum or spectrogram. The basilar membrane takes the complex wave delivered to it by a variety of acoustical and physical means and divides it into pressure patterns of differing frequencies.

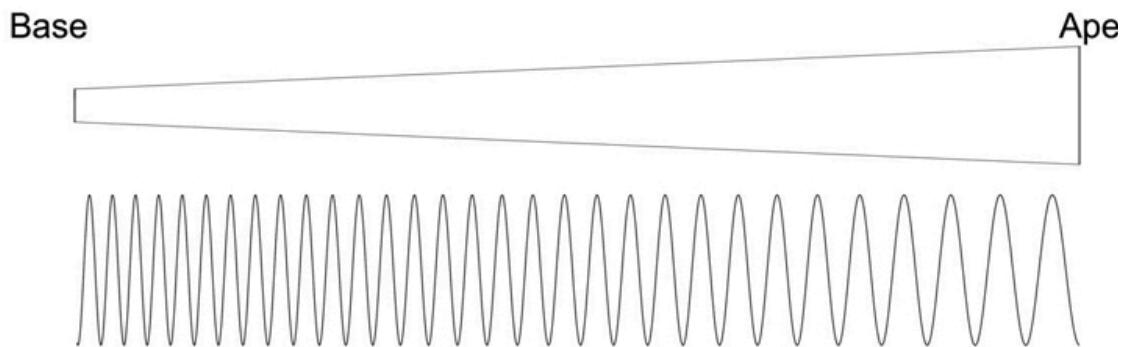


Figure 3.4 Schematic of uncoiled Basilar membrane. Frequency response is organized high to low from base to apex. Source: Created by author.

However, the basilar membrane is not an isolated structure, nor does it filter pressure patterns as a Fourier transform does (more on this in [chapter 6](#)).¹⁷ Adjacent to the basilar membrane (and within the cochlear duct) is a complex of structures called the *organ of Corti* which contains thousands of tiny hairlike cells called *stereocilia* (see again figure 3.3).

When a specific region on the basilar membrane moves in response to a pressure wave of a specific frequency, the adjacent stereocilia cells are bent. This triggers an electrical signal which travels through the *cochlear nerve* to the brain. The basilar membrane accomplishes frequency filtering by means of its nonuniform stiffness. At the base of the cochlea, where the stapes introduces vibration, the basilar membrane is stiff and narrow. Toward the apex of the cochlea, the basilar membrane is wider and more flexible.

It may not be intuitive to think of the response of a physical membrane in terms of its *resonant properties*, but the change in the stiffness of the basilar membrane along its length means that it is best able to move at fast rates of change where it is narrow and stiff (at the base). Slower rates of change find compliance closer to the apex, where the membrane is more flexible. Its resonant properties then change along its length. This means that a low-frequency pure tone co-presented with a high-frequency pure tone will literally stimulate two different physical locations along the basilar membrane.

However, the input to the cochlea is typically not a pure tone. The pressure pattern in the air is almost always a complex wave. In fact, it is typically a complex wave that is the combination of simultaneous complex sounds from multiple sources. As that complex wave propagates through the cochlea, different moments in the wave capture both the amplitude and rate of change in pressure. This suggests that as this wave travels within the cochlea—rising and falling in pressure at different rates over time—a given point on the basilar membrane will move when the rate of pressure change best matches its physical resonance. It is not so much that specific stereocilia are sensitive to specific frequencies. It is more that the basilar membrane is generally resistant to moving in response to any frequencies save for those that align with how flexible it is at a given point. If it can react, it does. If it reacts, it stimulates those stereocilia.

With a pure tone, the pattern of pressure change is simple and oscillates between high and low values. This generates a sustained resonance response in the corresponding region of the basilar membrane. A periodic, complex wave will repeatedly stimulate similar regions of the

basilar membrane, but this pattern of local stimulation likely changes throughout the period, as explored in this chapter. As a pressure wave progresses toward its frequency-related location along the basilar membrane, it gradually rises to a pressure maximum. Thereafter, the wave quickly stops physically displacing the membrane. This suggests that a strong stimulation of the basilar membrane in response to lower frequency pressure patterns may cast a *masking* shadow over the portion of the membrane that responds to a higher frequency stimulus.¹⁸ The lower frequency tone may partially or completely prevent one from hearing the higher frequency tone.

The basilar membrane does not have infinitely detailed frequency resolution. When an area of the basilar membrane moves in response to vibrations of a given frequency, the adjacent stereocilia will also respond to a lesser degree. This overlapping region is called a *critical band of hearing*. Any two separate, near-similar stimuli affecting adjacent physical regions of the basilar membrane may result in literal crossed wires as the brain is challenged to untangle which response came from which stimulus. This is the physiological basis for the phenomenon of *auditory roughness*, a “buzzy” percept, which is explored in [chapter 7](#).

Finally, the basilar membrane itself is organized almost logarithmically, much like pitch. This means that the critical bands of hearing encompass wider frequency ranges from the low-frequency apex to the high-frequency base. For now, keep in mind that as a complex pressure wave passes over the basilar membrane, the higher frequency energy in the wave is more likely to fill in critical bands of hearing while potentially being partially masked by lower frequency energy. In this manner, the inner ear limits the information that it passes along the cochlear nerve to allow the brain to establish the details of the auditory scene.

As complex as these models are, they do not approach the complexity of the actual process of hearing. Several key dynamics are left for later consideration. First, the brain receives electrical impulses from different lengths of the basilar membrane at slightly different times, and it receives the same pressure wave at different times in each ear. The brain extracts meaning and assembles the structure of the auditory scene from these

timing differences. Second, the cochlea is not a passive structure that transfers vibrations from the outer world into the brain. It actively reacts to stimulation, to our thoughts and intentions, and to our predictions informed by our experiences. It is, quite literally, powered by the brain, which regulates sensitivity, amplifies quieter sounds, and directs our attention. Finally, and maybe most importantly for musicians, the brain appears to respond to ear training. One can learn to *hear* with greater specificity and detail.

Notwithstanding that hearing is much more complex, a midlevel model of the ear lays the groundwork for many of the ideas that follow:

1. The ear receives complex stimuli and reshapes them by means of the varying resonances of the ear canal, ossicles, and basilar membrane. The effect of these filters varies with the intensity of the stimulus, and different regions of the basilar membrane respond to different frequencies.
2. The basilar membrane has resolution limitations. Any stimulus falling within the critical band of another stimulus will trigger auditory roughness. These critical bands are not *fixed* along the basilar membrane. They are a dynamic response to stimuli and can start and end anywhere along the membrane.
3. The resulting signal becomes a series of electrical impulses interpreted by the brain by means of pattern analysis.¹⁹

Taken together, these ideas prepare us for an end-to-end model for understanding the acoustical and psychoacoustical phenomenon of voice production. Such a model would begin with the response of the vocal tract

to an impulse and end with the processing of a complex pressure wave by the inner ear and brain.

IS SOUND A COMPLEX WAVE OR A SPECTRUM?

Taking in the above suggests the benefit of conceptualizing voice production in the physical world of pressure changes over time, rather than the mathematically abstract way that sound may be represented in a spectrum. Is it possible to synthesize a facsimile of a voice by building up a complex wave from sine tones? Yes. Is that how the human voice creates its sound? No. Human voicing operates in the world of complex pressure patterns up to the point of frequency filtering by the cochlea. This means that despite the value of spectral analysis, we must switch back to a time-and-pressure-domain model before we can imagine how the pressure pattern interacts with its environment and changes as it moves on to each step in the process. This also allows us to think of acoustical aspects of the voice as real, physical parts of the pressure and flow dynamics at work in the vocal tract. They are simultaneously a sound and a physical phenomenon.²⁰

Auditory transduction then can be understood as the process through which the complex vibrations that move the eardrum are turned into nerve impulses, which are in turn received and interpreted by the brain. This process changes aspects of the sound *before* it reaches the brain.²¹ If we consider that the mathematics of the Fourier transform will typically preferentially represent the repeated information within a periodic pattern in terms of harmonics of the slowest repeating pattern (the fundamental), perhaps we stand to reassess whether spectrograms represent voices in a manner consistent with our perception of those voices.²² In a spectrum we see the summary of a snapshot in time (or a series of such snapshots in a spectrogram) in this process of sound generation, propagation, and perception. In an actual living human ear,²³ this frequency division reduces information and introduces qualitative aspects of timbre that are not reflected in a spectrogram. If for no other reason than this, I believe it is crucial to evaluate whether our evolving models account for the contributions of the ear.

It is worth repeating: The frequency decomposition carried out by a Fourier transform may be conceptually matched in nature by the action of the cochlea, and applying a Fourier transform to the pressure wave at any step of the complex process of voice production—even if just at a model level—absolutely tells us something valuable. It is helpful to understand the differing spectra of the glottal impulses facilitated by the vocal folds as they thicken and thin, for example. The high-frequency energy more present in the former (and generally lacking in the latter) is perceptually relevant. This kind of contrast is reflected in a spectrogram far more intuitively than in a flow glottogram or waveform. *However, we must be cautious not to arbitrarily impose cochlea-like frequency filtering at this earlier step in the process.* The cochlea is not stimulated by the vocal fold source sound prior to the response of the vocal tract. We do not hear the vocal fold source sound. We hear the interaction of that source pressure impulse, the vocal tract, the room and/or electronic audio signal chain, and the mediation of auditory transduction as a functional whole. We must be cautious to recognize that any such image represents a way to *think* about aspects of that underlying reality. It does not depict actual separate harmonics that physically exist in space and through time, as though they were generated by discrete processes.

Thus, within the voice pedagogy and vocology literature, a spectrum always needs to be qualified. If the spectrum seeks to represent some aspect of the pressure wave prior to the information-destructive process of auditory transduction, we must recognize that it risks misrepresenting the physical nature of voice production. If a spectrum appears to represent the sound as we perceive it, we must recognize that the ear introduces qualitative aspects to the percept of the sound *after* the instant in time that is displayed in the spectrum.

INTRODUCING PITCH, AUDITORY ROUGHNESS, AND TONE COLOR

As I wrote above, a Fourier transform mathematically summarizes repeated aspects of an audio waveform and displays it as a spectrum. It does not illustrate the physical nature of voice production. It also does not display how we *perceive* that sound. With that established, we now confront what a spectrogram or power spectrum leaves out. **Pitch**, **auditory roughness**, and **tone color** characterize much of what is missing. These perceptual qualities will be unpacked individually in [chapters 5 through 9](#). [Chapter 10](#) offers a framework for reading these qualities into a spectrum or spectrogram.

In 2017 I proposed a framework for thinking about the objective aspects of timbre relevant to a voice.²⁴ This model explores the voice through the simplified intersection of the three perceptual qualities mentioned above: pitch, auditory roughness, and tone color. More specifically, these terms point to meaningful contrasts along continuums: *pitch to noise*, *pure to buzzy*, and the *vowel-like qualities* between dark and bright. I believe that as a singing voice migrates through pitch, intensity, timbre, and registration options, incredibly powerful aural cues for singers and voice teachers lie at the intersection of these percepts. Indeed, one may perceive the sound of a singer as being composed of multiple, simultaneously sounding, qualitatively separable percepts. The relative presence, absence, strength, and weakness of these qualities can characterize genre, registration, and affect. These qualities are quickly identifiable, easily differentiated, and *do not exist prior to the processing of the pressure wave by the cochlea*. This means that different sound sources may have timbral qualities in common simply because they stimulate the ear in a similar manner. More on this in [chapter 4](#).

This model allows voice teachers and students to begin to assign pedagogically relevant perceptual qualities to the spectrum and spectrogram images they see in their textbooks and journal articles. However, the more meaningful practical application may ultimately lie in

ear training and functional listening. One should be able to predict and discuss the sound (not simply the phoneme) captured in those images with reasonably high accuracy. One should also be able to understand, identify, label, and predict obligate perceptual changes in a voice as the singer changes pitch, intensity, timbre, and registration, and to do so without the help of a computer.

POTENTIAL CONCLUSIONS AND APPLICATIONS

Singing is, at its core, an aurally experienced art form. As with our other senses, the process of hearing limits and colors our experience of our surroundings. The teaching models found in voice pedagogy texts to explain voice acoustics are typically not based on how we hear. Instead, these models rely on Fourier transform–based computer measurements, predominantly of speech. While measurements of speech do tell us something true about the pressure wave as measured in the air, the resulting conceptual frameworks may miss qualitative aspects of perception relevant to the singing voice. By coming to understand how the ear and hearing brain work, we can ask how one might possibly perceive a singing voice. This is another way of asking what sounds a human voice is capable of making, which cuts to the heart of what it is to train a singer.

DISCUSSION QUESTIONS

- What is psychoacoustics, and what questions about the voice does it make actionable?
- What outside fields do voice pedagogy books draw on to explore voice acoustics? Can you give an example from a voice pedagogy textbook?
- What mathematical process is used to generate the spectra and spectrograms commonly found in voice acoustics chapters in voice pedagogy textbooks?
- What is the nature of the input of the vocal folds to the vocal tract? Does it generate harmonic frequencies or broadband frequency energy? Support your answer.
- Is the vocal tract a more efficient resonator when the vocal folds are open or closed?
- Is there a specific location of a given vocal tract resonance (or formant) in the vocal tract?
- Which parts of the ear help to reshape the spectrum of the incoming pressure wave?
- In what frequency regions is the ear most sensitive, and how does this vary with the intensity of the sound?
- Why does the middle ear need to step up the amplitude of the pressure change received by the eardrum?
- What property of the basilar membrane allows it to filter for various frequencies?

NOTES

1. Johan Sundberg, *The Science of the Singing Voice* (DeKalb: Northern Illinois University Press, 1987), 77.
2. C. Julian Chen, *Elements of Human Voice* (New Jersey: World Scientific, 2017), x–xi; Ingo R. Titze, “Generation and Propagation of Sound,” in *Principles of Voice Production* (Iowa City: National Center for Voice and Speech, 2000); Thomas J. Hixon, Gary Weismer, and Jeannette D. Hoit, “Acoustic Theory of Vowel Production,” in *Preclinical Speech Science: Anatomy, Physiology, Acoustics, and Perception*, 3rd ed. (San Diego: Plural, 2020), 289–326.
3. Titze, *Principles*, 151.
4. Eric J. Heller, “Mechanisms of Hearing,” in *Why You Hear What You Hear* (Princeton, NJ: Princeton University Press, 2013), 360.
5. Tsutomu Chiba and Masato Kajiyama, *The Vowel: Its Nature and Structure* (Tokyo: Phonetic Society of Japan, 1958), 149–54.
6. Brad Story, “The Vocal Tract in Singing,” in *The Oxford Handbook of Singing* (Oxford: Oxford University Press, 2019), 151–54.
7. Reinier Plomp, *The Intelligent Ear* (Mahwah, NJ: Lawrence Erlbaum Associates, 2002), 146.
8. Heller, *Why You Hear*, 415–28.
9. A representative list of voice pedagogy texts includes Kenneth Bozeman, *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013), 3–4; Oren Brown, “Resonance and Power,” in *Discover Your Voice: How to Develop Healthy Voice Habits* (San Diego: Singular, 1996), 80; Jean Callaghan, “Resonance,” in *Singing and Science: Body, Brain, and Voice* (Oxford: Compton, 2014), 73; Barbara M. Doscher, *The Functional Unity of the Singing Voice*, 2nd ed. (Lanham, MD: Scarecrow Press, 1994), 93, 139; Meribeth Bunch Dayme, “Resonation and Vocal Quality,” in *Dynamics of the Singing Voice*, 5th ed. (New York: Springer, 2008), 126–28; Wendy LeBorgne and Marci Rosenberg, “Resonance and Vocal Acoustics,” in *The Vocal Athlete*, 2nd ed. (San Diego: Plural Publishing, 2021), 96–98; Scott McCoy, *Your Voice: An Inside View*, 3rd ed. (Gahanna, OH: Inside View Press, 2019), 21–26, 38–51; James C. McKinney, *The Diagnosis and Correction of Vocal Faults* (Long Grove, IL: Waveland Press, 2005), 24; Donald G. Miller, *Resonance in Singing: Voice Building through Acoustic Feedback* (Princeton, NJ: Inside View Press, 2008), 13–28; Garyth Nair, *The Craft of Singing* (San Diego: Plural Publishing, 2007), 58–62, 183–86; James Platt and David M. Howard, “Applied Vocal Acoustics and Acoustic Registration,” in Janice M. Chapman and Ron Morris, *Singing and Teaching Singing: A Holistic Approach to Classical Voice*, 4th ed. (San Diego: Plural Publishing, 2021), 198; Kari Ragan, “A Systematic Approach to Resonance,” in *A Systematic Approach to Voice: The Art of Studio Application* (San Diego: Plural Publishing, 2020), 195–96; Brad Story, “The Vocal Tract in Singing,” in eds. Graham Welch, David M. Howard, and John Nix, *The Oxford Handbook of Singing* (Oxford: Oxford

University Press, 2019), 148; Johan Sundberg, *The Science of the Singing Voice* (DeKalb: Northern Illinois University Press, 1987), 20.

For more comprehensive, modern scientific models of voice production, see Christian T. Herbst, Coen P. H. Elemans, Isao T. Tokuda, Vasileios Chatzioannou, and Jan G. Švec, "Dynamic System Coupling in Voice Production," *Journal of Voice*, in press (February 1, 2023); Ingo R. Titze, "The Source-Filter Theory of Vowels," in *Principles of Voice Production*, 2nd printing (Iowa City: National Center for Voice and Speech, 2000), 149–84; Zhaoyan Zhang, "Mechanics of Human Voice Production and Control," in *Journal of the Acoustical Society of America* 140, no. 4 (October 2016).

10. Sundberg, *The Science*, 20.
11. Gunnar Fant, *Acoustic Theory of Speech Production* (The Hague: Mouton De Gruyer, 1970), 19; Philip Rubin and Eric Vatikiotis-Bateson, "Measuring and Modeling Speech Production," in eds. Steven L. Hopp, Michael J. Owren, and Christopher S. Evans, *Animal Acoustic Communication: Sound Analysis and Research Methods* (Berlin: Springer-Verlag, 1998), 252–55; Sundberg, *The Science*, 20.
12. *Voice Production: The Vibrating Larynx*, DVD, Janwillem van den Berg and William Vennard, 1960 (University Park: Pennsylvania State University, 2013).
13. Norman J. Lass and Charles M. Woodford, *Hearing Science Fundamentals* (St. Louis: Mosby Elsevier, 2007), 60; Hixon, Weismer, and Hoit, *Preclinical Speech Science*, 521.
14. Hixon, Weismer, and Hoit, *Preclinical Speech Science*, 521.
15. David M. Howard and Jamie Angus, *Acoustics and Psychoacoustics*, 5th ed. (New York: Routledge, 2017), 74.
16. Heller, Why You Hear 423.
17. Soumyajit Mandal, Serhii M. Zhak, and Rahul Sarpeshkar, "A Bio-Inspired Active Radio-Frequency Silicon Cochlea," in *IEEE Journal of Solid-State Circuits* 44, no. 6 (2009): 1814–28.
18. Lass and Woodford, *Hearing Science Fundamentals*, 133–34.
19. Howard and Angus, *Acoustics and Psychoacoustics*, 151–52.
20. Herbst et al., "Dynamic System Coupling," 6.
21. Heller, Why You Hear 417–23; Plomp, *The Intelligent Ear*, 13.
22. Heller, Why You Hear 422–25.
23. Jozef J. Zwislocki, *Auditory Sound Transmission: An Autobiographical Perspective*. (Hove: Psychology, 2013). Zwislocki's work is groundbreaking.
24. Ian Howell, "Necessary Roughness in the Voice Pedagogy Classroom: The Special Psychoacoustics of the Singing Voice," in *VOICEPrints* (May/June 2017): 4–7.

4

What Is Timbre?

Before diving into a narrow—if pedagogically useful—subset of perceptual phenomena associated with the timbre of a singing voice, it may be helpful to define the broad idea of *timbre* itself. The official definition of timbre, according to the American National Standards Institute (ANSI), is:

that attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar. . . . Timbre depends primarily upon the frequency spectrum, although it also depends upon the sound pressure and the temporal characteristics of the sound.¹

This definition is widely criticized by those who work in speech and music perception. Roy D. Patterson (2010) suggests that “you might expect the definition of timbre to tell you something about what timbre is, but all the definition tells you is that there are a few things that timbre is not.”² Bregman (1994) calls it a “wastebasket category” and suggests updating the definition to “we do not know how to define timbre, but it is not loudness and it is not pitch.”³ Robert Cogan, writing in 1969, echoes this long-standing issue: “Timbre, of all the parameters of music, is the one least considered. It lacks not only an adequate theory, but even an inadequate one.”⁴

WHAT DIFFERS? WHAT IS THE SAME?

What remains when pitch and loudness are controlled? It is not enough to label that which remains the timbre of a specific instrument, such as a soprano or violin. Or at least we must recognize that each of those instruments has a wide timbral palette. The ANSI definition leaves space to consider that changes in pitch and loudness impact the timbre of that soprano or that violin. As Kai Siedenburg and Stephen McAdams (2017) put it, "There does not exist *the* bassoon timbre, but rather *a* bassoon timbre at a given pitch and dynamic."⁵ Put another way, we anticipate that the timbre of that singer or instrument will change as their pitch and loudness change.

Complicating this further, the ANSI definition of timbre centers the *difference* between two hypothetical sounds. Pitch and loudness are identified as loci of potential similarity. Given the same pitch and loudness, what remains is timbre. Interesting questions arise if we turn the definition around to ask whether there are other *similar* qualities between otherwise differently timbred sounds.

This is an attractive question in a voice pedagogy context. For example, if a study claims that a classical soprano may benefit from aligning their lowest vocal tract resonance with their fundamental starting at D5, the teacher who can hear the effect of that adjustment across their entire population of differently timbred sopranos is advantaged. This remains true whether the teacher chooses to guide their singer toward or away from that acoustical alignment based on their aesthetic targets. If the strong fundamental at that pitch creates a consistent percept in a listener's ear—which is to say that one can learn to dependably *hear* it—this idea leaves the world of the subjective (or academic) and enters the world of the practical.

This is easy to explore with an experiment. Record a singer executing an arpeggio at a moderate tempo starting on the pitch G4 (see [figure 4.1](#)). Next, record a violin (or another treble instrument) executing the same passage. Ideally have the instruments match vibrato and intensity—

the underlying concept is more clearly illustrated the more similar the samples—but this is not strictly necessary. Take each sample, the voice and the instrument, and pass-filter the fundamental using a spectrograph like VoceVista Video Pro, Praat, or a similar program. This means you only let that part of the spectrum play back. Play the resulting simple sounds back-to-back and consider what they share in common. What is similar? See Lab #2.



Figure 4.1 G major arpeggio. Source: Created by author.

First, let us focus on a property of these sounds that Wayne Slawson (1985) suggests might be termed *sound color*: an inherent, qualitative aspect of a sound unaffected by how that sound evolves over time.⁶ Or consider how Reinier Plomp (1966) states that “in addition to pitch, simple tones have timbre . . . that have some resemblance, depending upon frequency, with particular speech vowels . . . [and] that only on the basis of this assumption the timbre of complex tones can be understood.”⁷

Sit with this idea for a moment.

By pass-filtering the fundamental from two separate sound sources, we have compared separately produced simple tones of similar frequencies. According to Plomp, the timbre of these tones will vary with frequency. If the frequencies of the two samples are the same, the tone colors will be the same. Put more directly, *changing the frequency changes the tone color* regardless of the source. Listen to the two samples back-to-back again. Notice that as pitch rises and falls in both the voice and treble instrument, the tone color changes respectively. In anthropomorphized terms, one might perceive that, in a manner consistent with an important aspect of the continuous change from [u] to [ɔ] on a sustained pitch, the

vowel-like color *opens* as pitch rises. One might suggest that these tone color changes are vowel-like. A non-anthropomorphized view might suggest that human vowels are tone color-like. This assertion suggests that speech leverages a set of deeper tone-color percepts. Figure 4.2 sequences this logical process.

For a simple tone, frequency determines timbre (tone color).



The fundamental extracted from two separate sound sources is a simple tone that changes tone color with frequency



The frequency-dependent tone color change is the same for both extracted fundamentals, regardless of sound source



The tone color contribution of the fundamental persists when reintegrated into the complex wave.



Therefore: Both the voice and treble instrument share a common tone color quality based on the contribution of the fundamental.

Figure 4.2 A logical flowchart illustrating similarities between different sound sources that are currently precluded by the official ANSI definition of timbre.

Two important insights flow from this demonstration. The first is that pitch and timbre appear to have a tighter relationship than the ANSI definition of timbre would suggest. *Controlling for pitch constrains and informs the possible resulting timbre.* The second, implicit in Plomp's observation, is that the tone color of that pass-filtered fundamental may be understood as one contributor to the timbre of the complete sound. This too is easily demonstrated experientially. Return to the original recordings of the voice and treble instrument. Prime the listener by first playing the pass-filtered sample followed by playing the full spectrum. If using the filtering feature of VoceVista Video Pro, fade in the rest of the spectrum while the sample loops playback. If using Praat, listen to the isolated fundamental in a Spectrum object and then listen to the entire spectrum. Observe that the tone color contribution of the fundamental is likely persistent in the context of the remainder of the spectrum. See Lab #3.

For many, this may be the first practical step in the acquisition of functional listening skills. The tone color aspects of the timbral percept associated with the fundamental displayed on the spectrum or spectrogram is both *similar* between the voice and the instrument and also *noticeably different* from the rest of each spectrum. Comparing and contrasting the timbre of the voice and instrument illuminates not just how they differ but also that which they have in common. Now, a leap: If you learn the functional coordination that is associated with the percept of that fundamental—which, I will argue much later in this book, relates to the transglottal airflow volume—you can literally *hear* the function in a singer. The voice teacher who is trained to hear these qualities objectively may not just link what a spectrogram shows onscreen to aspects of a singer's sound. They may quickly learn to identify these qualities without the aid of a computer.

TIME AND TIMBRE

The next several chapters will focus on the aspects of timbre that are conceptually similar to the tone colors of the fundamentals explored in figure 4.1. In a way of thinking, these properties exist *outside of time*. That is to say, as time passes, their intrinsic character does not change. They are time invariant. They are either present with some amount of loudness or they are not, and their role in the overall timbre may change over time.

By contrast, there are myriad ways in which other timbral qualities might unfold over time. This includes the nature of the onset or attack, the character of how a sound is sustained—is it steady or does it fluctuate?—and the way the sound is released. Further, one could consider that changes in loudness (that is to say, *dynamics*) are obviously tied to the passage of time. The spectral aspects of changes in loudness can be understood in terms of their relative presence or absence rather than a change in their intrinsic quality. Return to the pass-filtered fundamental from the soprano and the treble instrument. Notice that they likely differ in timbre in other noticeable ways, even as they share tone colors. See Lab #4.

As a voice teacher who spends time being present with and reacting to sound, I would never suggest that time-invariant qualities of timbre are more important than those qualities that unfold over time. The latter might be thought of as the signifiers that transmit easefulness, control, confidence, appropriate registration, and style. However, I do believe that conceptually separating these two ideas is an essential step toward clarifying functional listening.

As an example, if a singer moves into problematic pressed phonation—and pressed phonation is only a problem if it functionally inhibits singing—one might notice a decreased easefulness in their ability to smoothly change pitches. The pressing will attenuate the lower, warmer portion of the spectrum while favoring the brighter, buzzy part, but the sound of this miscoordination itself *must* unfold over time. The inability to smoothly

change pitches will be shown over time by virtue of the behavior of the warm, bright, and buzzy qualities. In this example, the warm/bright/buzzy qualities are time-invariant aspects of the sound. They are individually unrelated to the passage of much time.

POTENTIAL CONCLUSIONS AND APPLICATIONS

Is timbre complex? Yes. Any multidimensional system or phenomenon is complex. Does this complexity prevent one from parsing out dependably identifiable aspects of timbre from such complexity? I argue that it does not. This idea flows from the principle at the core of the preceding pages: The act of sound perception takes place in a physical world by means of physical systems that follow rules. The ear coequally determines the sound of a singing voice. In the text that follows, I will address aspects of perception one by one with a focus on pitch, auditory roughness, and tone color. These aspects will be grounded in explanations, scholarly context, and ear-training examples to enliven one's experience of hearing singing.

The big takeaway for the above material is that it matters how we think. Again, our conceptual models both serve and trap us. These models not only define just how we expect reality to proceed but also limit the questions we think may be appropriate to ask. Those interested in incorporating auditory perception into their work with voices—either in the studio or in an academic classroom—may benefit from bringing significant nuance to the ANSI definition of timbre. Understanding the principles that follow leads to the creation of newer conceptual models that illuminate that most ephemeral aspect of the singing voice: *the sound*.

DISCUSSION QUESTIONS

- What is the official ANSI definition of timbre? In what ways is this definition problematic?
- Discuss the idea that there may be loci of similarity beyond pitch, loudness, and duration between two otherwise differently timbred sounds.
- The idea that an aspect of a sound might exist outside of time is problematic, as all sounds exist in time. Please restate this idea in your own words, as intuitively as you can.
- Were you able to hear the tone color similarities between the voice and instrument when carrying out the experiments outlined in this chapter? Did you notice the between-instrument similarities and how the fundamental contrasted with the rest of each instrument's spectrum?
- Reflect on the assertion that “the ear coequally determines the sound of a singing voice” in the context of [chapter 3](#), which argued for an end-to-end model to understand voice production. Where is the sound of the voice *created*? Can this question be answered?

NOTES

1. ANSI, *Psychoacoustic Terminology: Timbre* (New York: American National Standards Institute, 1994).
2. Roy D. Patterson, Thomas C. Walters, Jessica Monaghan, and Étienne Gaudrain, "Reviewing the Definition of Timbre as It Pertains to the Perception of Speech and Musical Sounds," in eds. Enrique A Lopez-Poveda, Alan R Palmer, and Ray Meddis, *Neurophysiological Bases of Auditory Perception* (New York: Springer, 2010), 223.
3. Albert S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press, 1994), 92–93.
4. Robert Cogan, "Toward a Theory of Timbre: Verbal Timbre and Musical Line in Purcell, Sessions, and Stravinsky," in *Perspectives of New Music* 8, no. 1 (Autumn–Winter, 1969): 75.
5. Kai Siedenburg and Stephen McAdams, "Four Distinctions for the Auditory 'Wastebasket' of Timbre," in *Frontiers in Psychology* 8, article 1747 (2017), <https://doi.org/10.3389/fpsyg.2017.01747>.
6. Wayne Slawson, *Sound Color* (Berkeley: University of California Press, 1985), 20.
7. Reinier Plomp, "Experiments on Tone Perception" (PhD diss., Institute for Perception RVO-TNO, 1966), 132.

5

Pitch

There is perhaps no aspect of music more important than pitch. It is notoriously prescribed by composers and meaningfully recomposed by performers. It would be nearly impossible to move through a day without experiencing pitch, even if just in the call of a bird, the buzz of a fly's wings, or the beep of an alarm clock.

Yet what is pitch?

Many musicians are aware that pitch has a relationship to frequency, usually measured in repetitions of a pattern per second, known as hertz (Hz). Even nonmusicians are typically aware of the central idea captured in the official definition of pitch: "the auditory attribute of sound according to which sounds can be ordered on a scale from low to high."¹ Frequency is objectively measurable, while pitch is how the brain perceives frequency. In the voice pedagogy community, pitch is typically differentiated from frequency by its perceptual nature.

Scott McCoy (2019) offers a concise example of how this has been expressed in the voice pedagogy literature:

Frequency is an objective measurement of vibrations per second; pitch, however, is a subjective perception that can be influenced by factors ranging from vibrato to timbre and intensity.²

This definition tells us more about what frequency *is not*, rather than what pitch *is*.

I would like to draw your attention to three nuanced ideas that coexist in the idea of pitch:

- 1. Pitch is an organizing property of sound.** We may compare two periodic sounds and order them as lower or higher than each other along the pitch continuum. For example, C5 (an octave above middle C) is *higher* than C4 (middle C). C5 is also higher than B4 (a seventh above middle C), but by a smaller increment in the pitch space. Please note that few musical sounds are perfectly periodic, but quasi-periodicity is sufficient for pitch perception.
- 2. Some sounds are pitched, while others are noise.** Pitched sounds feature the periodic (regular) repetition of pressure changes. “Noise” in the scientific sense features aperiodic pressure changes. Consider a piano tone and a cymbal crash. The two are qualitatively different. The sound of a cymbal does not just lack pitch (periodicity); it *is* a kind of noise (aperiodicity). Some definitions of noise include a cultural judgment based on whether the sound is unwanted or a nuisance. This is important to know, but I will not use the term in this way.
- 3. The idea of pitch contains complex oppositions.** When we consider a given sound, we can understand that some qualitative aspects of that sound may convey pitch, while other qualitative aspects of that sound may not. Within any single pitched sound, a portion of the spectrum has the potential to elicit a pitch percept and a separate portion may not. If a

sound has pitch, the potential still exists for some portion of that sound to be pitch-less noise.

This chapter will explore the qualitative aspects of pitch and suggest ways to locate these complex qualities in a spectrum or spectrogram.

PITCH AND TIMBRE

Recall that the ANSI definition of timbre introduced at the beginning of [chapter 4](#) specifically controls for pitch, and the timbre of a complex sound can change while holding pitch constant. This may give the impression that pitch and timbre are completely separable. However, the relationship between pitch and timbre is complex, interdependent, and must be qualified.

Plomp (2002) writes, "For the extreme case of a sinusoidal tone without harmonics, the tone's frequency as its single variable determines pitch as well as timbre."³ This echoes the long-held assumption in the psychoacoustics literature that "simple tones have timbre" (see [chapter 8](#) for a more-thorough history).⁴ The fundamental frequency of a periodic pressure wave determines the frequency domain energy present in its spectrum. In turn, this frequency domain energy constrains the timbre that can be created.

For example, play the highest and lowest keys on the piano. Note that the former has a simple, bright, /i/-like timbre, while the latter has a rich, dark, and complex timbre. The former can never elicit the timbre of the latter because of the frequency of its fundamental. The bulk of the timbre of the spectrum generated by the lower fundamental does not exist within the sound produced by the high fundamental.

Locating this question within a voice perception framework similarly challenges the independence of pitch and timbre. Changing pitch while keeping the vocal tract stable may not change the sung phoneme in some cases, but changes in timbre do take place. Bozeman initially observed this phenomenon and has explored its pedagogic value in depth. He has called this "passive vowel modification" and, more recently, "passive vowel migration."⁵ Both of these terms point to changes in a singer's sound that occur as pitch rises, without changing the vocal tract. Pitch and timbre may then be thought of as semi-independent, semi-related aspects of sound. Some timbres are only possible within a prescribed pitch range. Some pitch ranges limit the potential timbre.

HOW DO WE PERCEIVE PITCH?

There are two theories that each partially account for the phenomenon of pitch perception: the *place theory* and the *timing* (or *temporal*) *theory*. The place theory assumes that the cochlea performs its frequency filtering on a periodic pressure pattern somewhat like a Fourier transform does, by breaking the complex pattern down into harmonic constituents, each of which *stimulates a different physical location along the basilar membrane* in the cochlea. By contrast, the timing theory assumes that it is the auditory cortex that extracts pitch from *the delay between successive transient peaks* in the pressure wave. In a voice, these successive transient peaks typically take place at the start of each period.

Both of these explanations carry limitations. The place theory does a poor job of explaining phenomena like the *missing fundamental*,⁶ where pitch perception is robust in the absence of energy at the frequency that is equivalent to the fundamental itself.⁷ The analog telephone system leveraged this phenomenon to convey the pitch percept whenever the fundamental was below 300 Hz. Music played back on small speakers similarly elicits the pitch of a wide range of sounds despite its poor low-frequency response. Or consider the unamplified voice of a bass singer singing a very low pitch in a large space. Very little of their fundamental will radiate to audience members. Not only is their lowest vocal tract resonance unlikely to be lower in frequency than around 270 Hz,⁸ their lowest frequency energy will be the least directional part of their sound.⁹ Since the *place* on the basilar membrane stimulated by the fundamental does not appear necessary for pitch perception, the place theory appears to be an incomplete explanation.

Likewise, one limitation of the timing theory is that the auditory cortex cannot extract accurate pitch information from a periodic sound whose period is shorter than 0.2 millisecond. This is equivalent to a fundamental ($1f_0$) above 5 kHz ($1,000 \text{ ms} / 0.2 \text{ ms} = 5,000 \text{ Hz}$). Because pitch accuracy is known to decrease above 5 kHz, any sense of pitch in this high-frequency range likely has a different explanation.¹⁰

I will leave an exhaustive discussion of the neuromechanical processes leading to the cognitive experience of pitch to readers who wish to explore the literature cited in the endnotes. For now, it is enough to understand that neither theory fully accounts for the pitch percept.

WHERE PITCH IS—AND IS NOT—IN A SPECTRUM

Let us reflect on two intuitive ways to imagine how to locate the pitch percept in a spectrum or spectrogram set to show harmonics. One way is to assume that the fundamental (the lowest harmonic) is the pitch while the other harmonics add other aspects of timbre. VoceVista Video Pro appears to support this notion by drawing its pitch-tracking line overtop the fundamental on the spectrogram. Another way to imagine how pitch appears in a spectrogram is to view everything as part of the pitch. As we will see, both options are wrong.

When viewed on a spectrum or spectrogram, pitch as a percept arises from the lower frequency range of the total spectrum. Depending on the source consulted, this may be as low as the lowest five harmonics ($1f_0$ – $5f_0$)¹¹ or, at most, the lowest eight ($1f_0$ – $8f_0$).¹² Perceptually, these lower harmonics are said to *resolve* into the pitch. Higher harmonics are termed *unresolved* and do not contribute to the pitch percept. Indeed, from about the ninth harmonic ($9f_0$) and higher, the energy of a human voice progressively contributes a pitch-less, bright, buzzy noise akin to the sound of a cricket. This noise is distinct from other forms of stochastic noise (for example, air turbulence or white noise), but it is conspicuously distinct from the pitch percept. A voice featuring strong energy above the eighth harmonic ($8f_0$) will typically have a strong and cutting noisy element. This may seem academic, but keep in mind that we are on the lookout for dependable, qualitative aspects of timbre. If we know that only part of the spectrum generates the pitch percept, we can place that in qualitative opposition to the part that does not. Crucially, this means we can dependably *hear* with specificity vocal functions that either do or do not generate that higher frequency energy.

It is worth exploring the nature of this noise for a moment. We just learned that unresolved harmonics do not necessarily sound like other, more familiar types of noise. They are more similar to the buzzing sound of crickets than a crash cymbal. When isolated, these unresolved harmonics may elicit a pitch percept (or even several simultaneous pitch

percepts) of their own. However, unresolved harmonics typically do not dependably generate the *correct* pitch percept.

Crossing from harmonic eight to nine ($8f_o$ – $9f_o$) does not immediately introduce this noise. As Plomp (2002) notes, "The unresolved higher harmonics may also contribute [to the pitch percept], but to a lesser extent." The frequency of the fundamental ($1f_o$) plays a role as well. Above about C5, the ninth harmonic ($9f_o$) will always fall above our 5 kHz pitch perception threshold, decreasing the likelihood that such energy can resolve into the pitch anyway. This > 5 kHz unresolved frequency information may be identified as a separate percept, in a perceptually bright space *above* and distinct from the pitch. See Lab #5.

Note that although one may visually identify the slower (lower frequency) and faster (higher frequency) aspects of a waveform, that a spectrum or spectrogram far more readily allows one to identify the waveform's resolved versus unresolved components. Remember that the objective correlate to pitch in the waveform is either the duration of the periodic pattern expressed in milliseconds or the frequency of its repetition expressed in cycles per second, or hertz (Hz). The physical aspects of that pattern that are unresolved from the pitch typically occur within that pattern. There is no easy way to visualize which ripples in a complex waveform will unfold over time (resonate) to represent unresolved information.

Think back to the model explored in [chapter 3](#). Per period of voicing, the resonances of the vocal tract are excited by the drop in transglottal airflow facilitated by the contacting of the folds. These resonances then quickly dampen until they are re-excited by the next vocal fold-contacting event, and the result of this complex event is captured as a function of time in a waveform. The slower ripples (lower frequency) visible in such a waveform typically oscillate longer per period than faster ripples (higher frequency) do; this is because the vocal tract preferentially dampens higher frequency oscillations.¹³ One developing conceptual framework aligns the resolved portion of the spectrum with the slower ripples that are allowed to oscillate to the end of the period, while the unresolved

portion of the spectrum may correspond to the faster ripples that shock into existence at the start of every period, only to rapidly dampen.

I suggest a middle path of sorts. It is incredibly useful to follow the literature and describe the threshold between resolved and unresolved portions of the spectrum in terms of harmonics of the fundamental. It asks just a little more nuance of us to keep in mind that harmonics characterize repeatedly occurring energy within a repeating pattern, not necessarily persistently oscillating pressure patterns. This opens the door to the potential value of a time-and-pressure-based approach in terms of understanding resolved versus unresolved spectral energy that will be explored in [chapter 6](#).

Recall also from [chapter 3](#) that as the complex pressure wave propagates through the cochlea, the basilar membrane cannot clearly resolve what a spectrum or spectrogram may display as higher harmonics. The information is either so tightly clustered along the basilar membrane—or is sufficiently aperiodic as a physical phenomenon—that it becomes impossible to extract the pitch. In a way of thinking, we should discard that detail in a spectrum anyway. Please keep in mind that little research has been carried out to explore these thresholds with sung sounds. More work remains to be done.

Notwithstanding that thought experiment, we must keep in mind that both the place theory and the timing theory are understood to be processes of the cochlea and brain. This means that pitch arises *after* the processing of the stimulus by the inner ear.¹⁴ So despite the declaration that frequency and periodicity are the physical correlates to pitch, pitch only ever exists in the brain. Crucially, the pitch percept is not the cognitive experience of interpreting pressure patterns as they exist in the air. It is the cognitive experience of that phenomenon *after* it has passed through the process of auditory transduction. Spectra and spectrograms show the decomposition of the pressure wave as it exists in the air.

WHEN IS A FUNDAMENTAL A FUNDAMENTAL?

This discussion complicates the use of the term *fundamental*. Here we may do well to differentiate between its use in the time domain and its use in the frequency domain. Periodic voicing almost always generates a complex pressure pattern that repeats at the rate at which the vocal folds contact. Whatever number of times per second that complex pattern repeats is the *fundamental frequency* of phonation. So if the vocal folds contact 440 times per second, the entire complex pattern will repeat 440 times per second. This may be expressed as 440 Hz, the modern pitch standard for A4. We can also consider the duration of each individual repetition of that complex pattern. Measured in milliseconds, this is called the *pitch period*. To convert from frequency (Hz) to duration (ms) for this pattern, divide 1,000 by the frequency (Hz). To derive Hz from ms, divide 1,000 by the period duration (ms). This means that a frequency of exactly 440 Hz will have a period duration of ~2.27 milliseconds. So *frequency* in this case points to the rate at which the complex pressure pattern repeats.

Once we apply the Fourier transform to that pattern, we will likely find higher frequency harmonic components that occur at integer multiples of that fundamental frequency. In that spectral domain, only one of those components will share the frequency of the vocal fold contacting events. This component is *also* called the fundamental and is notated $1f_o$ or sometimes f_o , with the integer understood as "1." Depending on the pitch and vowel being sung, this fundamental may also contribute a great deal of tone color and energy. A period (a single instance of the complex repeating pattern) contains all the timbral information of the vocal tract resonances. This includes both fast (high-frequency) and slow (low-frequency) changes in pressure, some of which resolve into the pitch.

RESTATING THE PERCEPTUAL QUALITIES OF PITCH

To recap, the portion of the sound shown on a spectrum or spectrogram as approximately the lowest eight harmonics *literally forms* the bulk of our perceptual experience of the pitch. These harmonics capture both slower and faster aspects of the complex pattern that repeats once per glottal cycle. In most cases, this percept sounds different from the contribution of the fundamental shown in a spectrum or spectrogram.

In almost all cases, the pitch percept of a voice may be thought of as the *tone color* (remember back to the pure quality of the fundamental in [chapter 4](#)) of that fundamental plus the tone color contribution(s) of the within-period pressure oscillations represented in a spectrum as these higher harmonics. Remove the fundamental (in a spectral sense) from a voice, and the pitch percept will still exist. The *color* of the sound will change but not the pitch. These two components of the pitch phenomenon are frequently separable perceptually, which is to say that in many cases the fundamental may be identified as a discrete percept apart from other portions of the spectrum.¹⁵ See Lab #6.

This means that the aspects of the complex pressure pattern shown as harmonics one through eight ($1f_o$ – $8f_o$) in a spectrum or spectrogram contribute *different* aspects of the same pitch percept, separable by *qualitative characteristics*. If we consider for a moment the idea that different frequency ranges elicit different frequency-dependent tone colors, the information in that lower portion of the spectrum will likely cover several different tone color ranges. The pitch percept will feature *multiple timbral aspects* that differ *qualitatively*. That is to say, the pitch percept is frequently made up of more than one tone color, and this integrated percept is separated in a spectrum. I will set aside the idea of frequency-dependent tone color for the rest of this chapter, but this idea will return in [chapter 8](#).

UNRESOLVED HARMONICS: THE NOISE IN PITCHED SOUND

Let us consolidate the limits of our ability to extract pitch information from the spectrum of a periodic sound. To review, the spectrum of a singer divides into lower resolved (pitch percept) and higher unresolved (generally buzzing noise) harmonics. The crossover takes place around the ninth harmonic ($9f_0$), but it is a gradual transition, and $9f_0$ should not be taken as a hard cutoff. Further, any information above 5 kHz is challenging to resolve into the pitch, regardless of where it occurs in the spectrum. These ideas point to a dependable percept typically associated with robust or high-pitched singing: Any sound with stronger upper harmonics or energy above 5 kHz will typically include a buzzing, noisy timbre. Regardless of whether this percept is strong, weak, absent, or co-presents with other strong percepts, when the spectrum does include such information, we hear it as buzzy noise. There are various physiological correlates to generating that higher frequency energy. I hope you are developing an intuition that high-frequency harmonics represent fast rates of change in the pressure wave, whether they are continuous or not. Therefore, this buzzy noise must correspond to some functional aspect of voice production that generates fast pressure changes. This will be explored further in [chapter 11](#).

At this point, I would like to illustrate these ideas with an example of a real singer. The complex waveform in [figure 5.1](#) has been transformed into an easier-to-read spectrum of harmonics (see Figure 5.2). Please see Appendix A for an orientation to these graphs. Neither of these images display the pitch as we perceive it. Figure 5.1 shows the frequency of repetition of the waveform pattern as a function of time (the duration of each repetition of the complex pattern in the waveform). Figure 5.2 displays this same information in a spectrum, which suggests that only the fundamental—the harmonic at the frequency equivalent to A_b4—aligns with the pitch; all the other harmonics visually align with higher values on the piano keyboard. The image would display the pitch more intuitively if

the lowest eight harmonics ($1f_0$ – $8f_0$) were grouped together (shown with brackets in figure 5.2). These lower harmonics are *resolved* into the pitch.

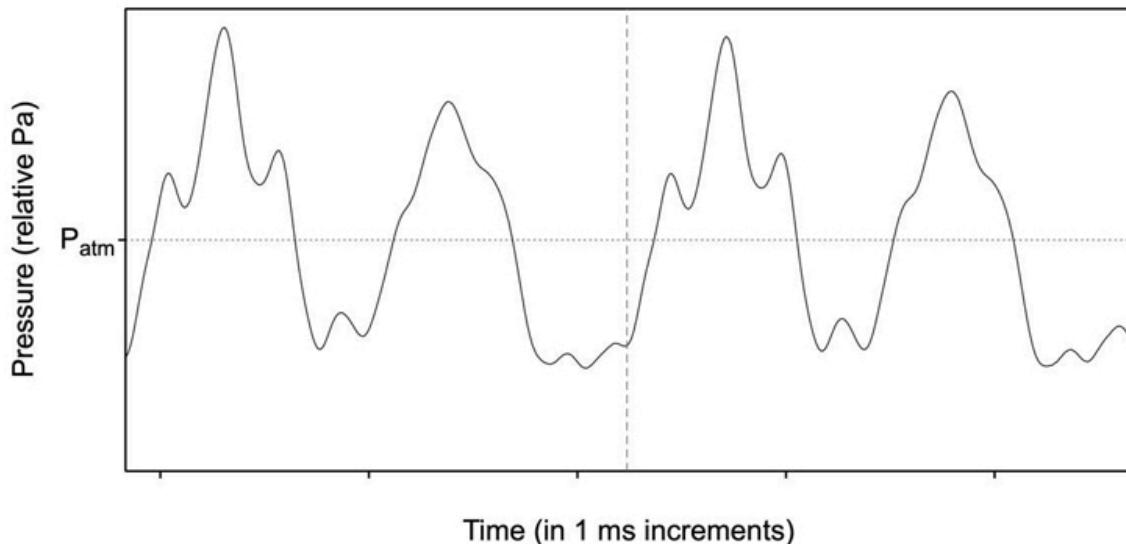


Figure 5.1 Two periods of audio waveform of a cisgender female soprano singing A♭4 [a]. Note the pattern repeats a second time starting at the dashed gray line. Source: Recorded by author under controlled conditions.

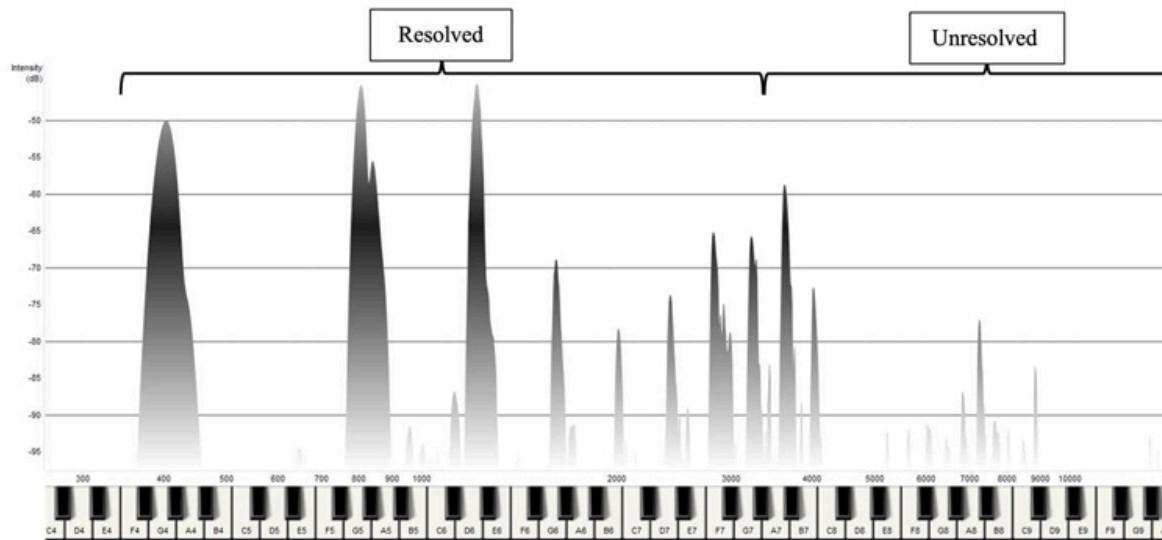


Figure 5.2 Power spectrum view A♭4 [a] from Figure 5.1. Each peak represents a harmonic that corresponds to the frequency equivalents of the pitches on the keyboard. The lowest eight harmonics have been bracketed together to show that they form the

pitch percept. Note that this percept arises after the processing of the inner ear. This percept does not exist in the compression wave that propagates through the air. Source: Recorded by author under controlled conditions.

ISSUES WITH THE SINGER'S FORMANT AND PITCH

The perceptual difference between resolved harmonics (pitch) and unresolved harmonics (pitch-lessness, or buzzy noise) points to an interesting paradox in a commonly accepted phenomenon. For the last half century, the singing community has largely accepted the idea that the singer's formant—sometimes singer's formant cluster (SFC)—is the reason the voice of an endogenous testosterone puberty classical singer can be heard over an orchestra.¹⁶ This bright and ringing quality is certainly strong in this kind of voice and frequently stands in for the aesthetic of classical singing, alongside other genre-defining technical choices such as a lowered larynx. Sundberg (2001) defines the SFC thus:

The “singer’s formant” is a prominent spectrum envelope peak near 3 kHz, typically found in voiced sounds produced by classical operatic singers. . . . It is mainly a resonatory phenomenon produced by a clustering of formants 3, 4, and 5.¹⁷

This has been demonstrated by means of masking recordings of singers with noise that imitates the theoretical long-term average spectrum of an orchestra.¹⁸ When this masking noise is applied to an endogenous testosterone puberty classical voice free of the SFC, the voice becomes hard to hear. When applied to a voice with the SFC present, that brighter portion of the singer's spectrum noticeably occupies a relatively “orchestra-free” frequency band in the spectrum. By contrast, treble voices—especially the voices of endogenous estrogen puberty classical sopranos—are said to not need or use the SFC, especially above the pitch C5.¹⁹ This exemption also applies to singers who use electronic amplification.

Setting aside the question of whether an orchestra ever produces a continuous sound equivalent to this masking noise—which suggests that we likely frequently hear the rest of the singer's spectrum at times—I would like to suggest that our understanding and appreciation of

brightness in all singing voices can benefit from more detail and specificity. Practicality drives this rationale, as *brightness* can present in many ways. Some ways of singing indeed have a strong SFC. Others feature a prominent harmonic lower than the SFC.²⁰ At other times, and especially in contemporary styles of singing, brightness is characterized by the contribution of harmonics higher than the SFC. If the SFC has a more objective technical definition (a specific frequency band amplified by a specific mechanical action), we ought to be able to attach some unambiguous perceptual aspect of the sound to the term. Otherwise, the practical pedagogic value of this term is problematically limited at best; we may as well revert to a general term like *brightness* instead.

Here is a specific example to tease out this issue. Given a SFC centered on 2,800 Hz, a tenor's SFC while singing a B_b4 will be represented on a spectrum by harmonics that are still low enough in the harmonic series to resolve into the pitch (see [figure 5.3](#)). If our tenor sings any pitch below about B3 (see [figure 5.4](#)), their SFC will likely be perceived as unresolved. It is still a bright and ringing sound, but an unresolved sound, nonetheless. This means the SFC frequency band increasingly contributes more buzzy noise than pitch as pitch drops. Were the SFC truly the only part of the voice that carries over an orchestra, we would hear the voice switch between pitched and pitch-less sounds based on the frequency of the fundamental and the frequency range of the SFC. This is obviously not how the voice behaves.

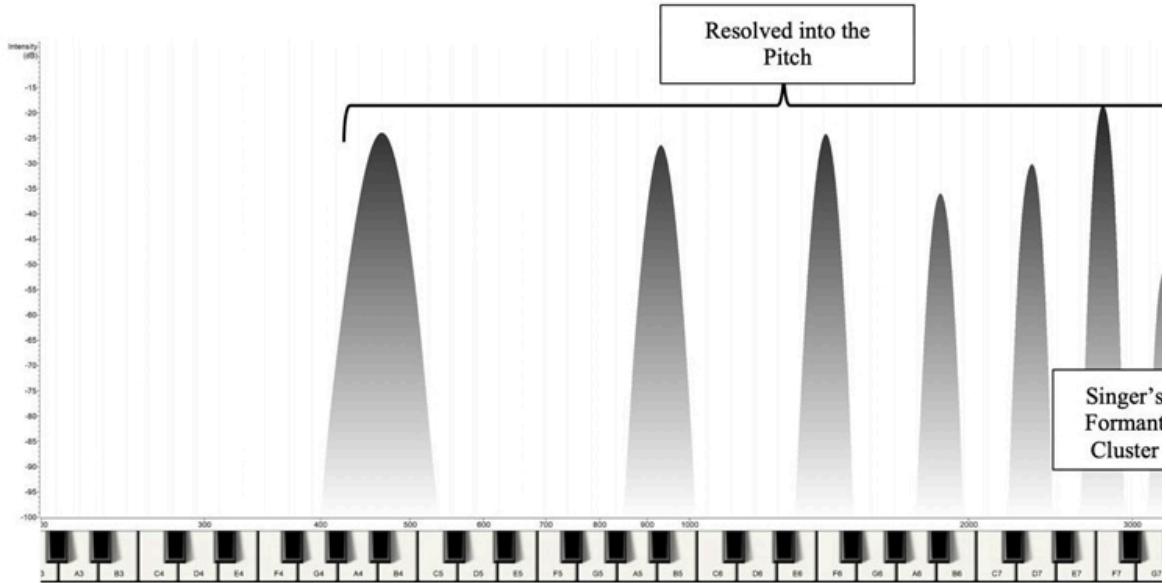


Figure 5.3 Schematic of a classical tenor singing B \flat 4. Note that at this fo the singer's formant cluster is resolved into the pitch. Source: Synthesized by author in Madde.

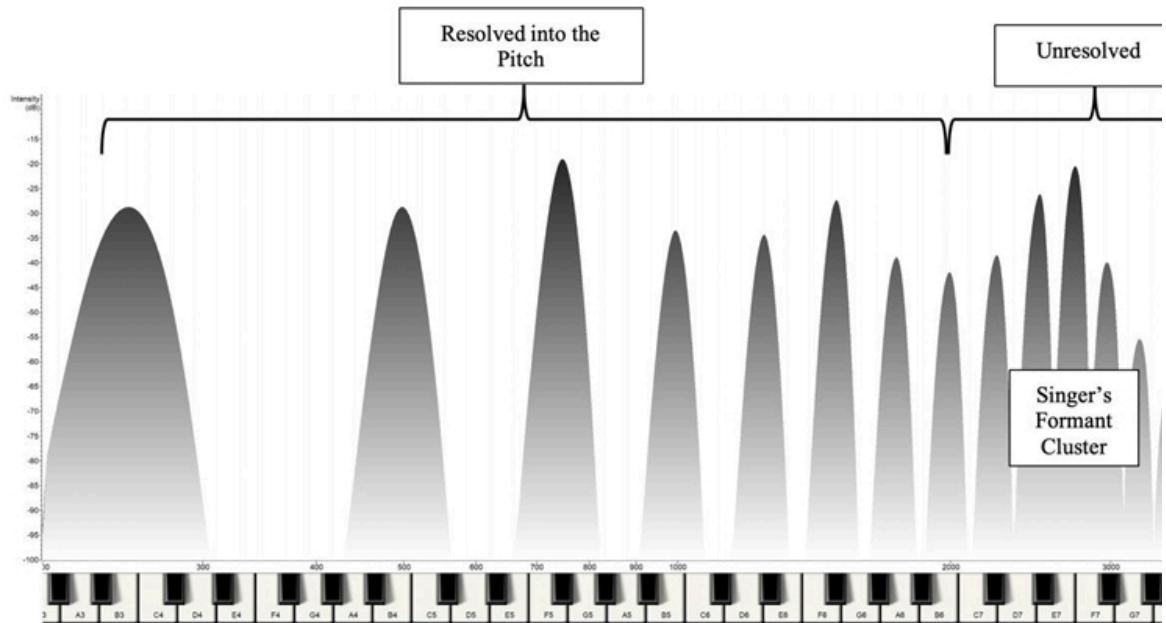


Figure 5.4 Schematic of a classical tenor singing B3. Note that the singer's formant cluster is unresolved from the pitch. Source: Synthesized by author in Madde.

At the writing of this book, I am unaware of studies that specifically explore the pitch resolution of harmonics in the singing voice from this

perspective. In particular, the sheer number of treble voices in the singing community would justify further research into brightness in singing at high(er) fundamentals.

Outside of the acoustics lab, this phenomenon is readily observed in situ, as shown in [figure 5.5](#). This figure models an A3 (top) and C♯4 (bottom) sung by Dmitri Hvorostovsky. The SFC of the C♯4 clearly resolves into the pitch, but the SFC of A3 takes on a noisy quality that challenges pitch identification. I speculate that identifiable transitions in classical singing voices may yet come to be associated with the transitions of the SFC from unresolved to resolved harmonics (rather than just pitch to pitch). Another way to say this is that as pitch rises, one would expect aspects of the singer's sound to transition from a pitch-less buzzy percept to a pitched buzzy percept to a pitched pure percept. This is further covered in [chapter 7](#).

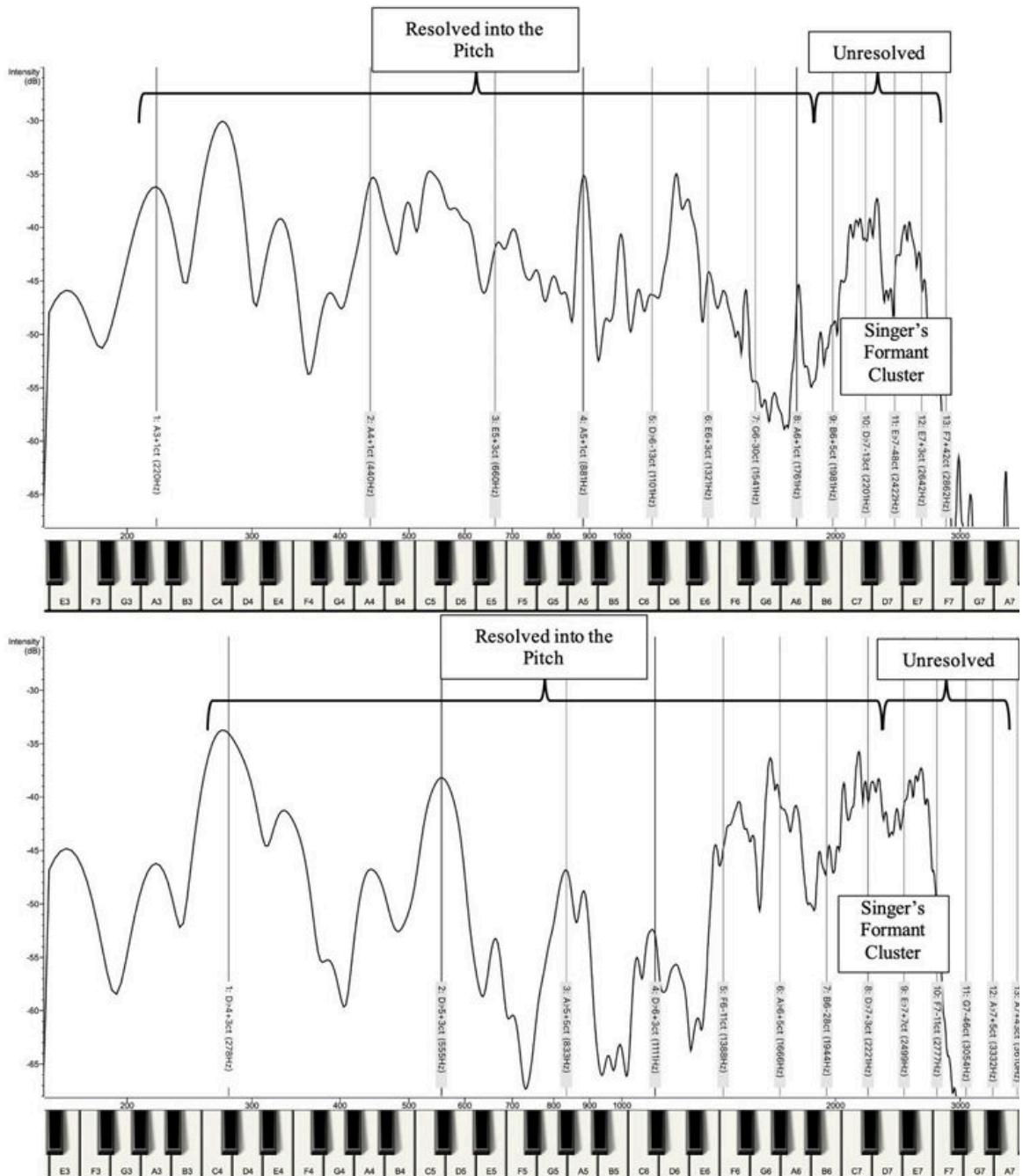


Figure 5.5 Long-term average spectrum (LTAS) of sample of A3 (top) and C \sharp 4 (bottom) sung by Dmitri Hvorostovsky. As this is a live recording with orchestra, the frequency regions aligning with Hvorostovsky's harmonics are indicated in each image. Note that the SFC for the A3 (top) begins at about the ninth harmonic and extends up to the thirteenth harmonic. The SFC for the C \sharp 4 (bottom) falls substantially within the range of resolved harmonics. Source: Met Live HD Broadcast of *Ernani* (Giuseppe Verdi), February 25, 2012, <https://www.youtube.com/watch?v=Mt7wYGysu0Q>.

Note that this is a commercial recording, and as such we must be cautious to assume we have heard exactly what this singer sounded like up close. However, this recording is sufficient for the purpose of demonstrating that his SFC *as heard on this recording* is not capable of transmitting the pitch percept.

BEGINNING TO UNDERSTAND PITCH IN OTHER VOICES

Momentarily setting aside the more thoroughly studied voices of classical basses, baritones, and tenors, let us consider the implications of pitch perception for the study of voices more broadly. In particular, let us consider the study of endogenous estrogen puberty classical singers, countertenors, and contemporary singers regardless of sex or hormonal regime.

As we have discussed, most speech research historically targeted formants between 300 Hz and 3.4 kHz. As such, many spectrographs default to a frequency display range that shows no information above 4 to 5 kHz. We recall that this condensed range makes sense in the historical context of the analogue telephone bandwidth and the study of speech and language. But the spectra of many ways of singing routinely exceed the upper frequency range of speech. Consider the same soprano from figure 5.1 singing an A♭ 4 (figure 5.6, above) and an A♭ 5 (figure 5.6, below). When the spectrograph caps the visible frequency range at 5 kHz, up to twelve harmonics are visible for the pitch A♭ 4. However, this truncation leaves only six harmonics visible for the A♭ 5. This suggests that no unresolved harmonics exist in this voice. At minimum, one would need to know to look for them in the frequency information above this graph. This should push us to ask whether the hidden frequency information is perceptually relevant.

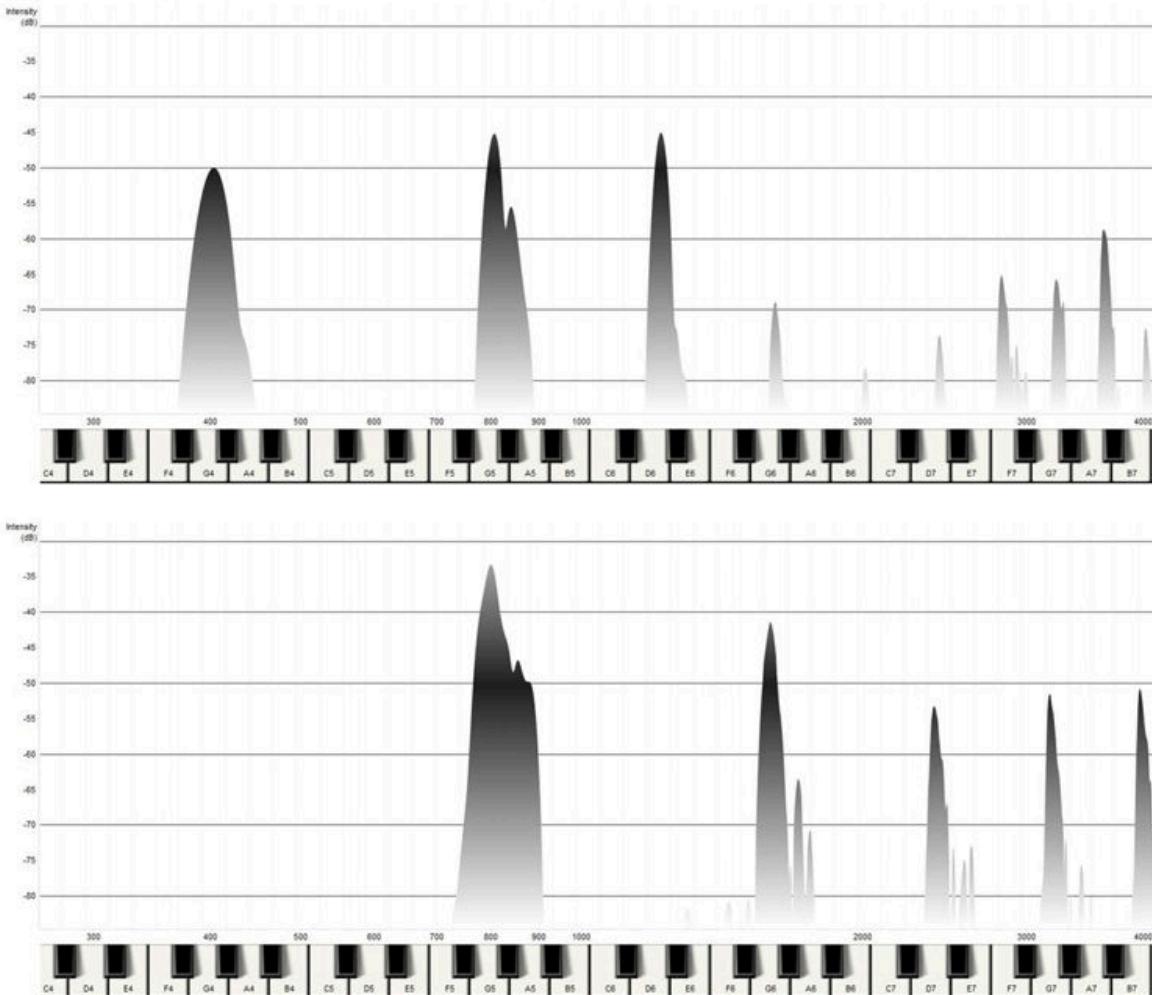


Figure 5.6 The soprano from Figure 5.1 singing an A_b4 (above) and A_b5 (below). Note that when the frequency range (horizontal axis) is capped at 5 kHz, only six harmonics are visible at the higher pitch. Source: Recorded by author under controlled conditions.

I must reiterate that the portion of the spectrum shown as unresolved harmonics will contribute something perceptually different than the portion shown as resolved harmonics. This is a qualitative difference relevant to the *sound* of a singing voice. In some singers there will be little spectral energy above 5 kHz. Consider the edge case of a low-intensity, speech-pitch range lullaby sung to a child on the verge of falling asleep. Most likely there will be no high-frequency energy; it almost certainly will not significantly impact language cognition even if it were perceptible. Even so, the artificial visual limit constrains the imagination. I

would go so far as to suggest that it inevitably limits the questions one may be likely to ask about how to listen to these voices.

By extension, a spectrum capped at 4 to 5 kHz implies that a voice singing above the treble staff will obligatorily elicit only a resolved sound. We know that is not necessarily so. It further suggests that the broad perceptual quality of *unresolved brightness* is only available to lower voices. The converse is also deeply problematic: The SFC model may push singers to imagine that sub-SFC energy may make no meaningful contribution to the sound in classical bass, baritone, and tenor voices. Additionally, the harmonics shown above the fundamental in [figure 5.6](#) (bottom) contribute to both the timbre *and the pitch percept*. One cannot assume that the timbral quality of a treble voice above the treble staff is fully captured in the sound of its fundamental, despite its obvious power. It is also worth pointing out that by comparison to their unamplified counterparts, singers who use electronic amplification may well capture sounds with significantly more of this high-frequency energy.

The voice pedagogy and vocology communities should grapple with the question of whether to disregard such high-frequency energy. Linguists suggest that we have captured enough of the spectrum in figure 5.6 to preserve language cognition. But that view excludes aspects of the sound that may be integral to the aesthetics of the singing voice. If the focus of our study were a violin, we wouldn't think to constrain spectrum in this manner, or to suggest that such a limited image reasonably captures the essence of the violin's character. In fact, raising the upper frequency limit of this image to 12 kHz reveals an additional, audible portion of the spectrum previously excluded (see [figure 5.7](#)). It is fair to debate the perceptual relevance of this higher frequency energy, but we have already learned that we can predict this spectral activity to contribute pitch-less noise. We cannot discuss that of which we are unaware.

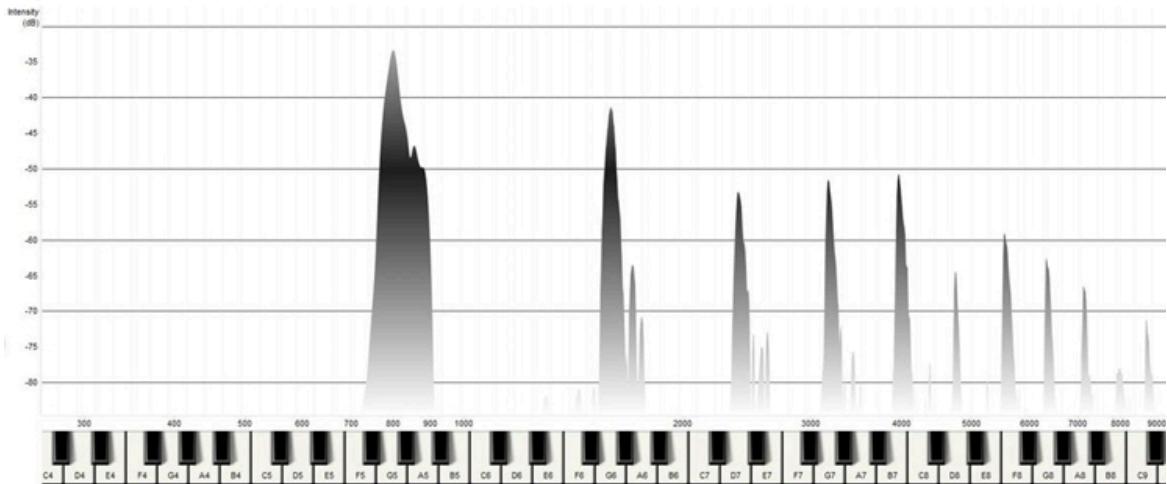


Figure 5.7 The soprano from Figure 5.6 singing an A♭5. Note that an additional spectral peak is revealed when the frequency range (horizontal axis) is extended to 12 kHz. Source: Recorded by author under controlled conditions.

Two additional examples highlight the distance between how we might model the acoustics of a voice in a way that excludes high-frequency energy and what the actual experience of that voice is. In “What Was I Made For?” from the soundtrack to the 2023 film *Barbie*, Billie Eilish sings with a distinctively breathy sound. Listen to the excerpted spectrum from the lyric “I” found around 0:11 in the recording (see figure 5.8). Four harmonics appear below 2.5 kHz, indicating periodic, pitched aspects of her sound. The peaks centered at 3.5 kHz and 10 kHz are dominated by breathy turbulence and convey little to no pitch information related to the fundamental. I would suggest that we label this higher frequency energy as stochastic and unorganized. It adds qualitative, bright aspects to the overall timbre but does not impact vowel or pitch identification. Worth considering as well is the fact that high-frequency hearing loss is a natural consequence of aging.²¹ While I can clearly hear the qualitative aspects of her voice up to around 12 kHz, I cannot perceive her energy above that. When I called my ten-year-old daughter into the room, she was able to clearly hear higher peaks. Researchers using their own senses to contemplate the timbre of a voice may inadvertently impose their own hearing limitations on the descriptive models they create. This also has

implications for audio signal chain considerations in perceptual studies, and for our understanding of the auditory perception of those who wear hearing aids.

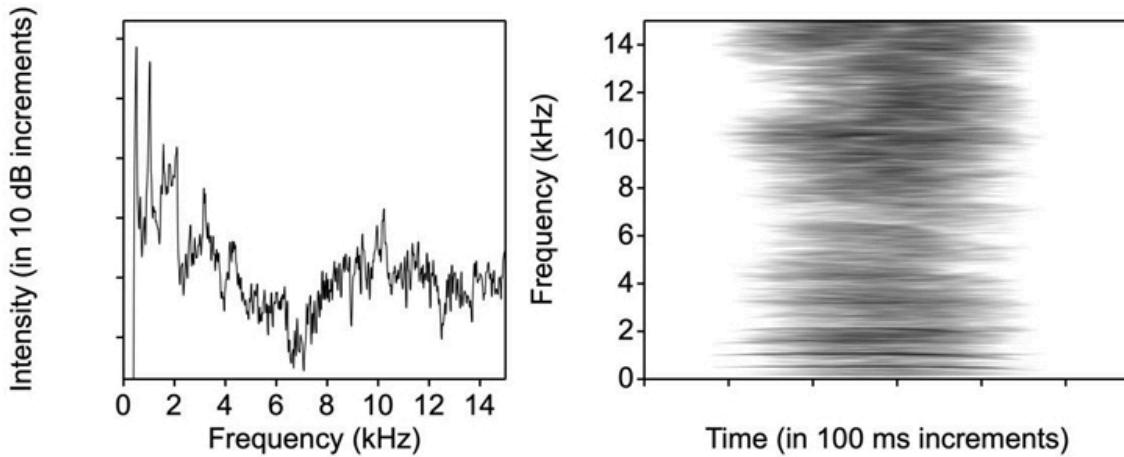


Figure 5.8 Spectrum (left) and spectrogram (right) of Billie Eilish singing “I” from “What Was I Made For?.” Source: <https://www.youtube.com/watch?v=cW8VLC9nnTo>.

Let us consider contrasting examples from James Brown’s 1965 single “I Got You (I Feel Good).” Mr. Brown’s initial fry scream with distortion, loosely transcribed as [wæu], contains little to no periodic oscillation of the vocal folds at all (see figure 5.9). Yet during the sustained [æ], we experience the chaotically excited, ongoing first resonance of the vocal tract (around 1,080 Hz) as the closest thing to a pitch. The higher frequency, equally noisy resonance peaks contribute nothing to the pitch percept. Instead, they add important qualitative elements to the overall timbre. A few seconds later, he sings “feel” (see figure 5.10) with reasonably clear harmonics in the spectrum up to 10 kHz, which is remarkable considering the age of the recording. I would argue that the now better-organized information centered at 7 kHz contributes a pitch-less, buzzy quality that distinguishes Mr. Brown’s sample from the turbulent noise in Ms. Eilish’s sample.

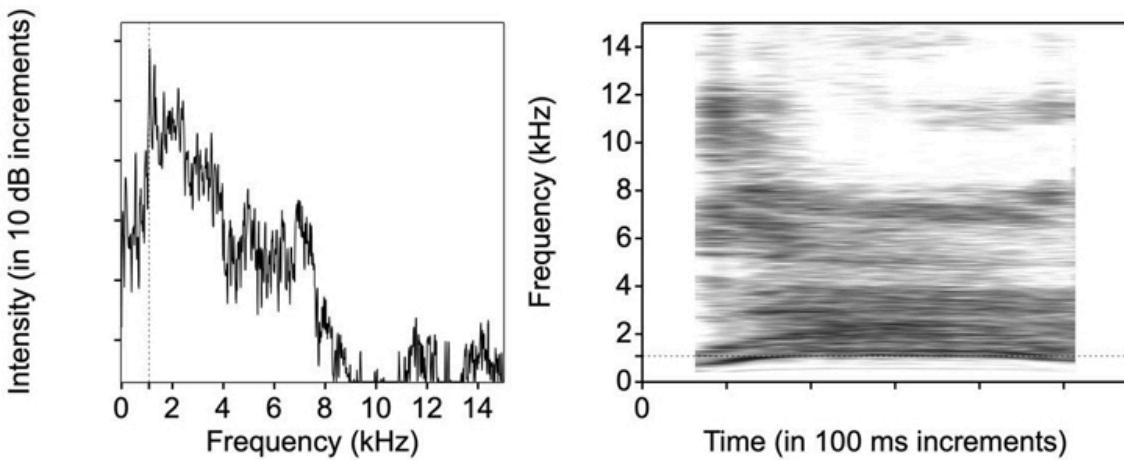


Figure 5.9 Spectrum (left) and spectrogram (right) of James Brown’s fry scream with distortion [wæu] from “I Got You (I Feel Good).” Aproximate pitch percepct at ~1,080 Hz indicated with a dotted line.
Source: https://www.youtube.com/watch?v=W-rn7i_ETYc.

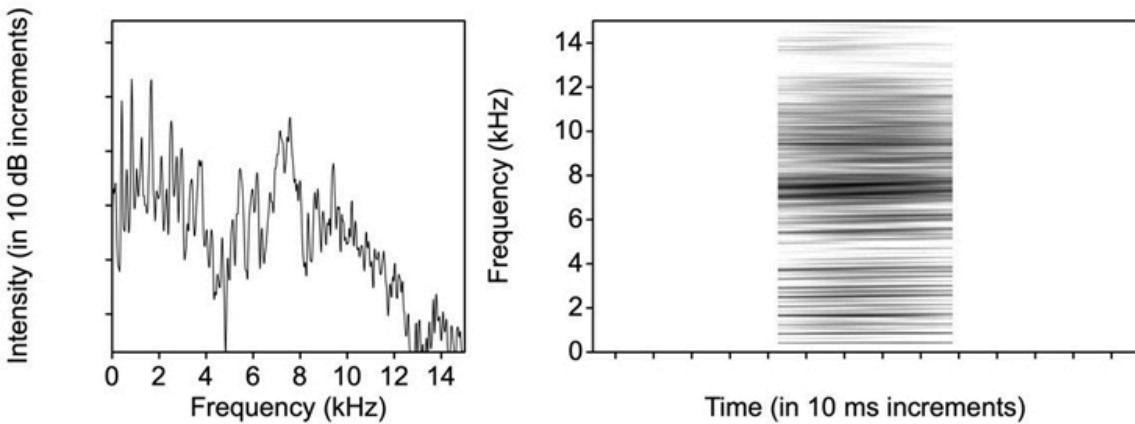


Figure 5.10 Spectrum (left) and spectrogram (right) of James Brown’s singing “feel” from “I Got You (I Feel Good).” Source: https://www.youtube.com/watch?v=W-rn7i_ETYc.

Some may critique my use of commercially recorded singers here. To be sure, the recording process has a mediating effect on the sound of a singer. But consider that we have long understood the electronic signal chain to be an integral extension of modern musical instruments. Were we to study the sound of an electric guitar, we would not require the player to remove the amplification or associated electronics they use to shape

their sound. One does not drill down to the essence of Jimi Hendrix's guitar playing by removing the distortion his amplifier provided. Artists like Hendrix crafted their sound to *include* the transformation of the original acoustic information by these electronics. Removing them destroys the ecological validity of such a study. In the case of Ms. Eilish's recording, she certainly benefits from her intimate proximity to the microphone and the electronic processes that help to balance and amplify the breathy and pitched aspects of her singing.

This should be an *a priori* assumption for those of us who study unamplified classical singers as well. It is considered the best practice to control the acoustical variables inherent in a concert hall by recording singers with laboratory equipment in a controlled setting. But does this not similarly risk harming ecological validity in potentially unsolvable ways? Perhaps we can append Kai Siedenburg and Stephen McAdams's statement that "there does not exist *the* bassoon timbre, but rather *a* bassoon timbre at a given pitch and dynamic"²² in a specific space at a specific distance with or without an electronic signal chain for a listener of a specific age and hearing acuity. Remember that radiation is as important a subsystem of the voice as perception. When our well-intentioned lab setups control variables in ways that destroy the timbral qualities inherent to a genre, we must pause to reflect whether we are measuring that genre anymore. At best this muddies the way forward, but we should not discount aspects of a singer's sound just because they were consciously facilitated by external means.

As we move forward to chapters that focus on auditory roughness, tone color, and the ways in which these time-invariant aspects of timbre interact in predictable patterns to explain familiar phenomena, I will similarly push to reform some widely accepted models to better represent the way in which humans actually perceive the sound of a singer.

CONCLUSIONS AND POTENTIAL APPLICATIONS

So, what is pitch? Pitch is many things. Pitch is a way to order periodic sounds, and changes along this continuum are a potent tool in coding expression into music. Pitch exists in opposition to noise. And pitch can be thought of as a way to parse a single sound into lower and higher frequency ranges that elicit qualitatively opposable percepts. At the very least, I encourage the reader to notice that the idea of pitch is complex, which complicates its utility as a control criterion in the definition of timbre. If pitch is considered a common feature of two otherwise differently timbred sounds, how do we account for the various ways in which pitch and noise may manifest within a single sound? We can then move beyond our initial definition that pitch is how humans perceive frequency to appreciate that this only tells us that pitch is not frequency. This does not actually articulate what pitch *is*.

Understanding pitch in this broad manner allows one to parse out parts of the spectrum that are *not* pitch but are rather contrasting, unresolved noise. We can observe, for example, that basic qualities in a voice change with pitch by observing how spectral regions of pitch and *pitch-lessness* intersect with the SFC. At a low fundamental, the SFC itself may in fact qualify as noise. At a high enough fundamental, the SFC may become part of the pitch percept. Pitch may also be generated by the periodicity of a resonance, independent of the behavior of the vocal folds. See the exploration of formant tuning and overtone singing in [chapter 6](#) for a more thorough exploration of this idea.

This framework also encourages one to extend the default upper frequency limit of 4 to 5 kHz set by most spectrographs. I argue that it is possible to find meaningful spectral information above this preset range, especially at high pitches. Because we have learned to characterize the perceptual distinction between the pitched and pitch-less portions of the spectrum, we may now begin to train our ears to *hear* whether the spectrogram is leaving information off the top of what it displays.

This basic idea of pitch versus pitch-lessness is necessary in part because it sketches out a basic scaffold to account for the opposing qualities that may exist within a single vocal tone. Noticing this foundational contrast is an important first step. Real insight as a voice teacher comes from attaching this perceptual quality to the functional adjustments of the singer, which will be explored in greater detail in [chapters 11](#) and [12](#).

In the coming chapters I will elaborate on this basic oppositional framework. But noticing the contrast is an important first step. The way in which the voice pedagogy and vocology literature typically presents the idea of pitch risks limiting our potential for intimately understanding the phenomenon of a singing voice. For this reason, the ways in which we address the perceptual qualities of auditory roughness and tone color will undergo similar consideration.

DISCUSSION QUESTIONS

- What is the official definition of pitch?
- Despite being controlled for in the official definition of timbre, pitch and timbre seem to be interdependent. A change in pitch may force a change in timbre. Please explain this idea with examples.
- Neither the place theory nor the temporal theory fully explains the percept of pitch. Please give an example where each theory fails to explain observable phenomena.
- If you remove the fundamental of a pitched sound but leave the rest of its harmonics, the timing theory suggests the persistence of pitch.
 - Would we expect the period duration of the waveform to change? Why or why not?
 - What aspects of timbre might we expect to change with the removal of the spectral fundamental?
- Where is pitch shown on a spectrogram or spectrum?
- A single tone has a pitch aspect and a pitch-less aspect. What underlying mechanisms might be contributing to this?
- Even given just what you have learned so far, discuss how pitched vs pitch-less qualities may be useful pedagogic guideposts when listening to a singer:
 - How might this framework shape our expectations for certain sounds?
 - How might this framework influence our understanding of the aural cues of certain voice types or certain types of registration?
- Which limits in research methodologies may have inadvertently misled the field of voice pedagogy in its search to understand the nature of high-pitched singing, and by extension, the high-frequency energy generally present in singing voices?

NOTES

1. Neil M. McLachlan, "Timbre, Pitch, and Music," *Oxford Handbooks*, published June 2016, DOI: 10.1093/oxfordhb/9780199935345.013.44.
2. Scott McCoy, *Your Voice: An Inside View*, 3rd ed. (Delaware, OH: Inside View Press, 2019), 36.
3. Reinier Plomp, *The Intelligent Ear: On the Nature of Sound Perception* (London: Lawrence Erlbaum, 2002), 23–24.
4. Reinier Plomp, "Experiments on Tone Perception" (PhD diss., Institute for Perception RVO-TNO, 1966), 132.
5. Kenneth Bozeman, "Vowel Migration and Modification," *VOICEPrints* 16, no. 2 (November–December 2018); Bozeman discusses passive vowel migration. Kenneth Bozeman, "Vowel Perception, Modification, and Motivation" in *Kinesthetic Voice Pedagogy 2: Motivating Acoustic Efficiency* (Gahanna, OH: Inside View Press, 2021).
6. Hermann L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th ed. (1877), translated by Alexander J. Ellis (New York: Longmans, Green, and Co., 1912), 153. The missing fundamental phenomenon has been well documented since before Helmholtz, who called them *differential tones*.
7. Kyoga Lee, "Pitch Perception: Place Theory, Temporal Theory, and Beyond," EE 391 Special Report (Autumn 2004), https://ccrma.stanford.edu/~kglee/pubs/klee_ee391_fall04.pdf, accessed 29 June 2021.
8. McCoy, *Your Voice*, 42. McCoy offers an average frequency for the lowest male vocal tract resonance.
9. Johan Sundberg, *The Science of the Singing Voice* (DeKalb: Northern Illinois University Press, 1987), 123.
10. Eric J. Heller, "Mechanisms of Hearing," in *Why You Hear What You Hear* (Princeton, NJ: Princeton University Press, 2013), 474.
11. David M. Howard and Jamie Angus, *Acoustics and Psychoacoustics*, 5th ed. (New York: Routledge, 2017), 143.
12. Sam Norman-Haignere, Nancy Kanwisher, and Josh H. McDermott, "Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex," in *The Journal of Neuroscience* 33, no. 50 (December 11, 2013): 19,454.
13. C. Julian Chen, *Elements of Human Voice* (New Jersey: World Scientific, 2017), 93.
14. Howard and Angus, *Acoustics and Psychoacoustics*, 137. The authors outline two competing theories for pitch perception: the place theory and the temporal theory. Neither theory completely explains pitch perception, but both locate the emergence of pitch between the inner ear and the brain.

15. Ian Howell, "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color" (DMA diss., New England Conservatory of Music, 2016), 41. The author labels this the *obvious true fundamental*.
16. Johan Sundberg, "Articulatory Interpretation of the 'Singing Formant,'" *Journal of the Acoustical Society of America* 55, no. 4 (1974): 838–44; Johan Sundberg, "Research on the Singing Voice in Retrospect," *TMH-QPSR, KTH* 45, no. 1 (2003): 11–22.
17. Johan Sundberg, "Level and Center Frequency of the Singer's Formant," *Journal of Voice* 15, no. 2 (2001), [https://doi.org/10.1016/S0892-1997\(01\)00019-4](https://doi.org/10.1016/S0892-1997(01)00019-4).
18. Sundberg, "Level and Center," 1.
19. McCoy, *Your Voice*, 49–50.
20. Donald Miller, *Resonance in Singing* (Princeton, NJ: Inside View Press, 2008), 4. As may be expected from cisgender male singers employing a third harmonic/second vocal tract resonance strategy.
21. Norman J. Lass and Charles M. Woodford, *Hearing Science Fundamentals* (St. Louis, MO: Mosby Elsevier, 2007), 119.
22. Siedenburg and McAdams, "Auditory 'Wastebasket.'"

6

Time-and-Pressure Domain Considerations of Pitch and Timbre

Any periodic sound can be explained as a spectrum of discrete frequency components even though it is actually a complex, repeating pressure pattern. This is a tension that exists within our pedagogic models. You may recall that while [chapter 3](#) foregrounds the utility of understanding the voice outside the constraints of a Fourier transform, the examples in [chapter 5](#) are illustrated using Fourier transform-based spectra and spectrograms. In other words, while [chapter 3](#) discussed the physical (time, flow, and pressure) reality of voice production, [chapter 5](#) frequently used the language of a harmonic spectrum. Sometimes a spectrum or spectrogram set to display harmonics best illustrates a concept. At other times a waveform brings more clarity. Respectively, these two approaches correspond to the *steady-state* or *harmonic theory* of voice production and the *transient* or *inharmonic theory* of voice production.

Both approaches are useful ways to model aspects of the same physical phenomenon of voice production. Both exclude information. The overlapping history of these acoustical models is long and fascinating and is tied to developments in mathematics, signal processing, and digital computers.¹ A full exploration of this subject would require a dedicated

book. Nevertheless, I would like to explore a few key concepts in this chapter to encourage fluidity.

Directly put, a *transient* is something “temporary, brief, [and] fleeting.”² In acoustics, this refers to “a sound of very short duration, e.g., the sharp attack (‘onset transient’) at the beginning of a note or any rapid change in sound level.”³ More specifically in voice science, transient refers to both uttered sounds with a sudden onset and short duration (for example, the consonants /t/ or /k/) and also the sharp changes in supraglottal pressure associated with the drop in transglottal airflow as the folds come into contact each glottal cycle. The transient theory then concerns the repeated interactions of transglottal airflow patterns and the resulting excitations of the vocal tract. These interactions are typically modeled using a waveform that graphs changes in air pressure over time or a spectrum or spectrogram set to show these glottal excitation patterns.

Steady state describes a system in which a pattern of vibrations is assumed to be continuous and periodic; that is to say, perfectly repeating.⁴ This means that although the complex pressure pattern generated by the repeated excitation of the vocal tract may contain multiple slow and fast oscillations that may or may not mathematically align perfectly with the fundamental, the vocal folds themselves periodically repeat at a meaningfully regular fundamental frequency. This allows one to characterize those faster, within-period oscillations in terms of the repetition of the frequency of that large-scale vocal fold oscillation. For this approach to have meaning, one must simultaneously consider multiple repetitions of the transient excitation of the vocal tract. The steady-state theory therefore concerns the mathematical averaging of multiple glottal cycles at once and is typically modeled using a spectrum or spectrogram set to display harmonics.

These two conceptual lenses do not inherently conflict. As Fletcher (1929) writes:

The difference in the two theories is not, as some suppose, a difference in the conception of what is going on while the vowel sounds are being produced, but in the method of representing or

describing the motions in definite physical terms. The . . . [transient] point of view enables one to visualize in a more direct way what is taking place and consequently is of greater value to the phonetician interested in the mechanism of speech production. It probably enables one better to grasp the fundamental characteristic differences between the vowels.

The . . . [steady-state] point of view is probably more useful to the engineer who is interested in designing telephone systems to properly transmit speech. The separation of the speech into its component frequencies makes it possible to see quickly which frequencies must be transmitted by the system to completely carry all the characteristics of speech.⁵

One of these theories characterizes the cyclical process itself, while the other summarizes the result of that process.

In normal voicing, it is the sudden drop in transglottal airflow upon the contacting of the vocal folds that initiates the impulse that excites the air in the vocal tract. The resonant response of the vocal tract dampens in a complex manner over a short amount of time. In these terms, we can understand voicing as the overlapping of the end of each resonant excitation of the vocal tract with the start of the next. Ultimately, it is the repetition of these patterns that creates the radiated pressure wave. The driving forces for this process are complex and feature important nonlinear concepts,⁶ but this simplification is nevertheless accurate at a model level.

Even so, it is also true that *if* this process is periodic (or in the case of a human voice, *quasiperiodic*), a Fourier transform is a useful tool for characterizing these repeating patterns. A Fourier transform allows one to average together long durations of time (relative to the timescale of a pitch period), which effectively summarizes key features that are present within the repeated pattern. Both lenses may be used when explaining voice production, but only one of them describes its physical nature.

In this chapter we will use the transient model to explore several aspects of timbre and pitch through a waveform. It is my hope that presenting this complementary conceptual model will both further

illuminate the complex nature of the physical stimulus received by the brain and provide some context for what a spectrum or spectrogram does and does not show.

MODELING VOICE PRODUCTION WITHOUT THE FOURIER TRANSFORM

What does it mean to understand voice production in the time-and-pressure domain without the help of a Fourier transform? Consider the effect of popping a balloon in a cathedral. The pop itself induces a sudden change in air pressure in and around the space formerly occupied by the balloon. That pressure change propagates into the surrounding air mass. A complex cascade of reflections and refractions emerges in the surrounding space as that wavefront bounces off walls and the open air between objects. Depending on the nature of the space, this process may set up longer lasting oscillations at some frequencies while others find themselves quickly damped. The interference and damping behavior of these patterns as the air mass eventually returns to atmospheric pressure *is* the reverberant quality of that space.

A similar process takes place more rapidly and on a much smaller scale inside the vocal tract. The contacting of the vocal folds facilitates a sudden drop in supraglottal pressure. This is the functional equivalent of our balloon pop. This drop in pressure swiftly propagates through the vocal tract, setting the air mass contained therein into motion. The resulting complex flow and pressure patterns encounter and interact with the boundaries between more-open and more-closed regions of the vocal tract, bouncing off its walls, openings, and changes in diameter in ways that cause a cascading series of overlapping responses to the initial impulse. This *is* the reverberant quality of the vocal tract, which characterizes the timbral character of its shape at the time of phonation.

At this point our analogy begins to break down. In a large room, the strongest patterns of reflection—the standing waves that fit the dimensions of the space well—oscillate slower than our lower pitch threshold of 20 Hz. We do not typically hear the rhythmic flutter reverberation of an acoustically live, large to medium-sized room as a pitch. Since the vocal tract is significantly smaller, its standing waves oscillate fast enough to hear as pitches. This means that the standing

waves of the vocal tract have pitch equivalents. Also, compared to a much harder surface like the stone of a cathedral, the energy in a soft(er), fleshy vocal tract dampens quickly, falling to a fraction of its original amplitude in mere milliseconds. Conversely, our stone cathedral can reverberate for six seconds or longer.

But the singing voice is not a single impulse and response of the vocal tract. It is a rapid series of such impulses and responses, each interrupting the previous one according to the timing of the glottal contacting events and the amplitude of the resonance pattern still moving over time in the vocal tract. Understanding this concept gives us the opportunity to reconceive a resonance of the vocal tract as a pressure pattern that changes over time as it oscillates and is damped, rather than as either a potential amplifier or a spectral peak. The faster (higher frequency) and slower (lower frequency) resonant oscillations within the vocal tract all lie atop one another in the waveform much as faster ocean waves ride atop the slower tides.

[Figure 6.1](#) shows the initial impulse of a single glottal contacting event and the subsequent excitation and rapid damping of a cisgender man's [ɑ]-shaped vocal tract response. This pattern has no periodic repetition and resonates here for around 35 ms. If somehow he could periodically repeat this every 35 ms, this would correspond to a pitch equivalent to about 28 Hz. But this pattern is not repeated here. It is an isolated event for the purpose of this illustration. Therefore, a harmonic series cannot describe this pressure wave any more than it could reconstruct the waveform of the attack-and-decay of a cymbal crash. The closest thing to a repeated pattern here (and it does not perfectly repeat because of both the damping and its own complex nature) is the slowest ripple (resonant response) of the vocal tract to that impulse.

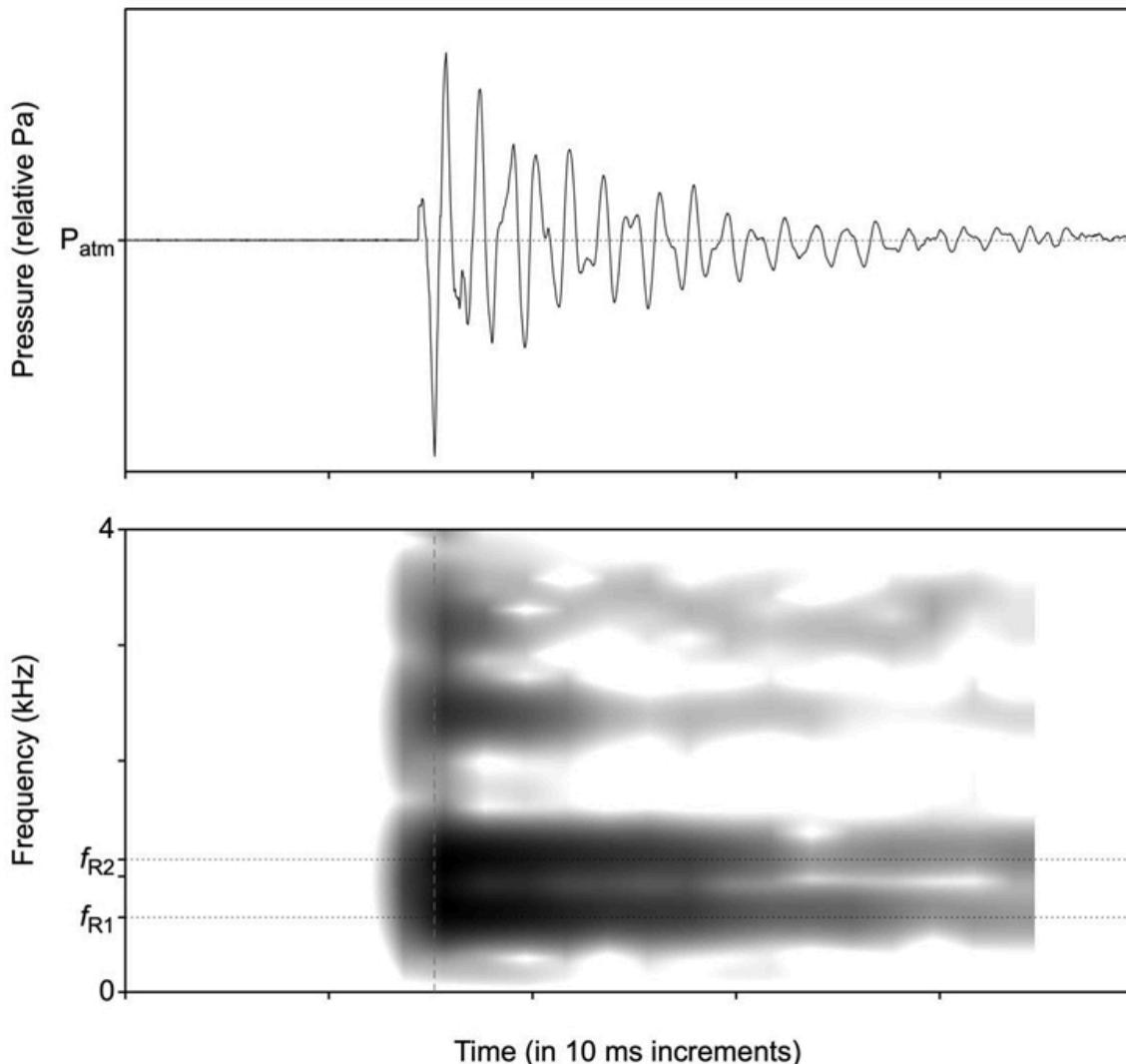


Figure 6.1 A single glottal impulse and decay tail of a cisgender baritone's [ɑ]-shaped vocal tract. f_{R1} and f_{R2} are indicated with horizontal lines. Higher frequency resonances are visible. Initial broadband impulse indicated with a dashed vertical line. Source: Recorded by author under controlled conditions.

The spectrogram shows several frequency regions (the strongest around 630 Hz, with smaller oscillations at 1,100 Hz, 2,400 Hz, 3,100 Hz, etc.) as preferential bands of acoustic energy that last longer than other bands. This waveform is not periodic in the way that we have observed previous waveforms of voicing. In fact, if the next impulse did not follow,

this is how all glottal impulses would likely play out in time. It is more like that single balloon pop in our cathedral rather than a series of them.

So, what are harmonics of a voice? At the timescale of a single period, when the timbres of the vocal tract resonances are created, there has not yet been a periodic repetition. *If harmonics require repetition, and the resonant response of the vocal tract takes place per period of voicing, resonances precede harmonics as a physical phenomenon.* Another way to think about this is that for all practical purposes, *the vocal tract resonances are the phenomenon that is repeated to create pitch in our brain and harmonics in a spectrum.* Harmonics then are a way to mathematically characterize even faster repeating patterns within the greater repeating pattern of the fundamental frequency of vocal fold oscillation. Harmonics do not necessarily point to persistent aspects of the resonant response of the vocal tract, or to continuous input from the vocal folds.

THE TIMING OF TIMBRE

We have set the groundwork for an interesting way to understand the separability of pitch and timbre in the human voice. We know that the pitch percept is generated when a periodic pressure pattern repeatedly stimulates the ear. We also know that these patterns themselves contain even faster oscillations that define important aspects of the timbre. Some of these aspects of timbre need to unfold over a certain amount of time. For example, let us consider an [u] vowel characterized by a strong first resonance (f_{R1}) around 324 Hz (~E4).⁷ As one attempts to sing pitches higher than E4 with this vocal tract shape, the vocal folds oscillate increasingly *faster* than the timbre of that resonance. Or, framed in the time domain, each cycle of the 324 Hz resonance takes 3.09 ms to unfold. Any pitch period shorter than this will interrupt the oscillation of these resonances, attenuating their timbral contribution.

This suggests that we may better understand the nature of vocal tract resonances if we think of them as fast or slow oscillations, rather than as high or low ones. The subsequent contacting event of the vocal folds per glottal cycle defines the duration of each period, and each resulting impulse *interferes* with the vocal tract's resonant response to the previous one. This means that pitch informs which timbres are available. *This suggests that the potential timbre for any pitch technically exists within the impulse and resonant response of a single glottal cycle.* It is the duration of the pattern itself that determines how far through the resonant response of the vocal tract the pitch period progresses.

If I can essentialize the above, the most obvious implication of this way of viewing voice production is that the repetition of the glottal impulse may disallow the existence of any color or timbre lower in frequency (slower oscillations with a longer period) than the fundamental. Despite the start of each period theoretically contributing a broadband impulse, a resonance with a waveform longer than $1f_0$ does not have time to unfold. Put another way, *some aspects of timbre have duration*; they must develop—or not—withing a single pitch period. At least within normal

periodic phonation, if that duration is not met, they functionally do not exist.

This may seem like a thought experiment rather than actionable information. However, it does begin to explain why timbre appears to migrate as pitch rises, regardless of whether one modifies the shape of the vocal tract. The more the increasingly shorter pitch period affects the unfolding pressure patterns in the vocal tract, the more pronounced this timbral migration. It is possible to sing pitches high enough that the resonant response of the vocal tract is interrupted nearly immediately, even before the initial amplitude of the impulse significantly decreases. This phenomenon not only has a precedent in historical vocal pedagogies⁸ but also forms the basis of vowel migration within Bozeman's framework of acoustic registers.⁹

This may be challenging to conceptualize on a first pass, so we may benefit from an example. [Figure 6.2](#) (top) shows the impulse and decay response pattern from figure 6.1 superimposed upon itself several times with a delay of about 7.5 ms ($1,000 / 7.5 \text{ ms} = \sim 133 \text{ Hz}$). Note that adding these pressure waves together yields a periodic pattern that looks credibly human (figure 6.2, bottom). Each "new" period combines with the end of the previous period. If the two periods are either both positive or both negative at any given moment, the overall amplitude increases. If they differ (for example, one positive and the other negative), they diminish the amplitude. In this example the overlapping ripples align well, boosting the amplitude.

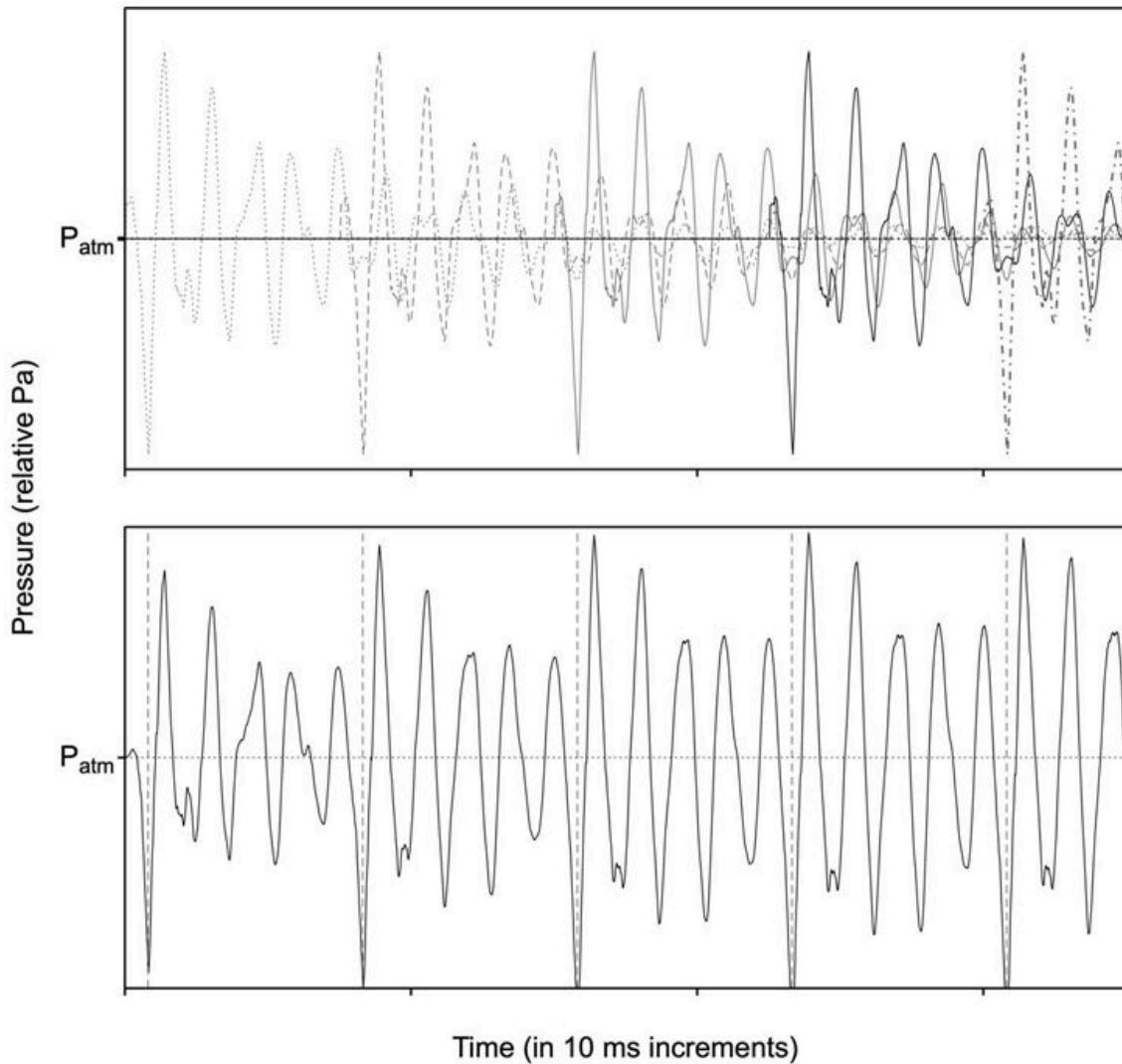


Figure 6.2 (Top) the impulse and decay tail from Figure 6.1 superimposed on itself four more times (dotted gray, dashed gray, solid gray, solid black, dashed-dotted black) with a 7.5 ms delay (~133 Hz, C3) as though the next vocal tract response interrupts and interferes with the decay tail of the previous vocal tract response. Adding these waves (bottom) generates a composite waveform that looks similar to a human voice. Source: Recorded and synthesized by author under controlled conditions.

Consider that the pattern in figure 6.2 (bottom) appears to preserve a meaningful amount of the decaying resonance of the vocal tract. The resonant frequencies captured within each period have time to oscillate

five times before their interruption by the next glottal impulse. Figure 6.3 illustrates the much higher pitch A6 (~ 1.71 kHz with a period of 0.58 ms) sung by a cisgender female soprano. Note the lack of much complexity within the pitch period save for a slight skew left per period. The resulting spectrum has a prominent $1f_0$ and a steep roll-off in the intensity of higher harmonics (see [figure 6.3](#)).

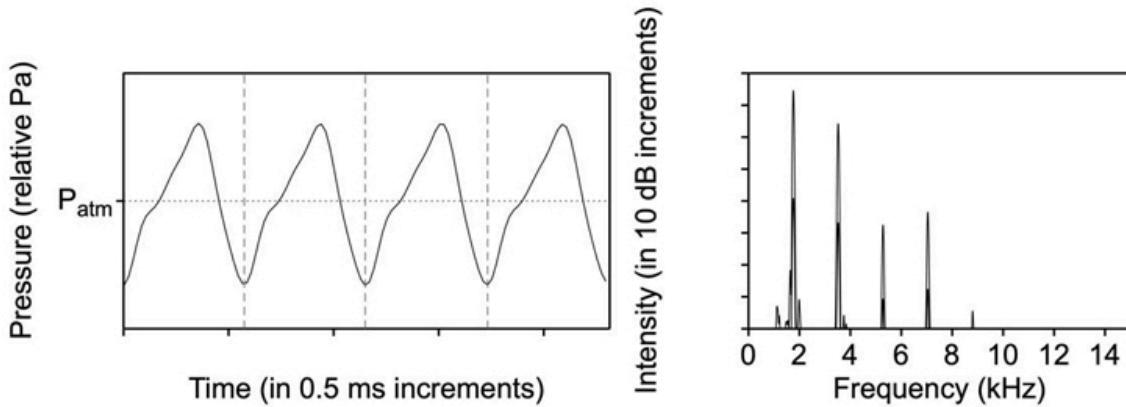


Figure 6.3 Four periods of the waveform (left) and spectrum (right) of a cisgender female soprano singing the pitch A6 (~ 1.71 kHz with a period of 0.53 ms). Approximate start of each period marked with a dashed gray line. Recorded by author under controlled conditions.

I propose a thought experiment. Consider that the pressure wave radiated by the soprano in figure 6.3 is *simpler* than the pattern in figure 6.2. The lower pitch allows multiple wide swings in pressure of that baritone's vocal tract resonances per period, while the soprano's waveform only contains one. Why is this? Does this soprano's vocal tract allow for fewer resonances than that baritone's does? The same soprano singing a C5 (~ 523 Hz) [a] (see [figure 6.4](#)) reveals a complex waveform and full spectrum, as one would expect from any singer.

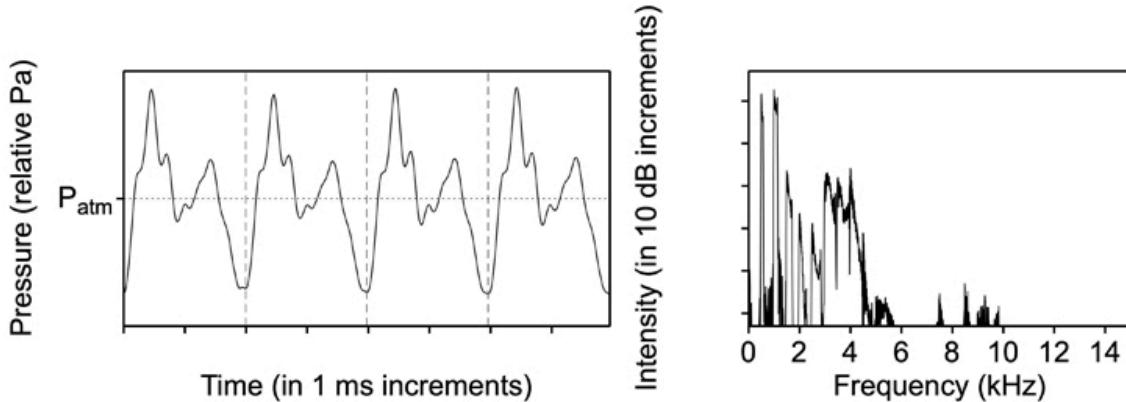


Figure 6.4 Four periods of a waveform (left) and spectrum (right) of a cisgender female soprano singing a C5 (~523 Hz with a 1.98 ms period) [a]. Approximate start of each period marked with a dashed gray line. Source: Recorded by author under controlled conditions.

What is the primary difference between these two pressure waves? It is the duration of each period. *At a very high pitch, the vocal tract does not have time to resonate the way it does at a lower pitch.* This is because resonances take time; they have a minimum duration as they move from low pressure through atmospheric to high pressure, and back through atmospheric to their starting low-pressure value once per cycle. It takes time to complete this cycle, and the speed at which this occurs determines their frequency. The resulting timbres that characterize low sung tones and high sung tones are wildly different because some aspects of timbre must develop over time. In this way, it is correct to assert that timbre has duration.

What about larger scale aspects of timbre that only occur when the spectrum fluctuates over time (recall the idea of time-variant aspects of timbre from [chapter 4](#)). From our new perspective, we may understand that timbral information is bound up in the pulse of resonances per glottal cycle (the way pressure patterns behave within each pitch period). At the same time, we may understand the way in which that information may change over time—the way these patterns vary (even subtly) from pitch period to pitch period. I would like to suggest that the fact that one might parse out this difference at all suggests the value in bringing nuance to the intersection of time and timbre.

WHAT IS FORMANT TUNING?

Formant tuning typically refers to the beneficial alignment of a given vocal tract resonance with a voice source harmonic to generate a pronounced peak in the spectrum, a *formant*. This is usually one of the two slowest vocal tract resonances (f_{R1} or f_{R2}) with a lower harmonic (most often between $1f_0$ and $4f_0$). The concept of formant tuning may have been most explicitly framed in this way by Donald Miller (2006),¹⁰ but it also appears conceptually in the work of Berton Coffin (2002)¹¹ and Kenneth Bozeman (2013)¹² and is similarly described by Titze (2000).¹³ According to the source-filter model, one might imagine the interaction of a discrete harmonic—presumably generated by the vocal folds—and a discrete resonance of the vocal tract. It is interesting to reconsider this framing if harmonics are more of a mathematical summary of the repeated excitation of the vocal tract than physical phenomena emanating from the vocal folds.

Up to this point, I have not made a meaningful distinction between the terms *formant* and *resonance*. What these words point to in a physical sense depends on the model being used to explore voice production. Within the source-filter model, one frequently differentiates a resonance from a formant in the following terms: A resonance is a potential amplifying power of the vocal tract, while a formant is the realized spectral peak once that resonance amplifies a vocal fold source harmonic.¹⁴ Within this model, that distinction makes a good deal of sense.

In the context of the transient theory model, I have been using the term *resonance* in a slightly different manner. Resonance in this case refers to the way in which a mass (in this case, the air in the vocal tract) continues to oscillate on its own once excited by an outside force. While one could certainly consider the *potential* oscillatory behavior of that air mass, we do not need to separately label that idea within the transient theory. Some may take issue with this framing; others may insist that the

term *formant* better captures this idea. Whatever label you prefer, just know which physical aspect of voice production this term points to here.

Thinking of this behavior of the singing voice in the time-and-pressure domain, we return to the question of how much the resonant response of the vocal tract has damped prior to the next supraglottal impulse. Remember, if the energy in the vocal tract is still oscillating at high amplitudes above the vocal folds when the next vocal fold contacting occurs, that carried-over energy may either productively facilitate or destructively interfere with the drop in supraglottal pressure brought on by the contacting of the vocal folds. This means that what one observes as formant tuning in a spectrum is unlikely to be the alignment of a discrete harmonic with a resonance in a physical sense. Instead, formant tuning may be more deeply understood as the alignment of one of the many repeating pressure drops of a resonance of the vocal tract with a subsequent supraglottal impulse. This can only occur when the frequency of the vocal tract resonance is at or near an integer multiple of $1f_o$.

In these terms, formant tuning may be more accurately labeled *resonance tuning*. This is true if (in terms of the steady-state model), the relationship is $1f_o$ aligned with f_{R1} (Bozeman's *whoop*¹⁵ or D. Miller's *hoot*¹⁶), $2f_o$ aligned with f_{R1} (Bozeman or D. Miller's *yell*¹⁷), or $3f_o$ or $4f_o$ aligned with f_{R2} (D. Miller's second formant tuning strategy).¹⁸ The fundamental ($1f_o$) aligned with f_{R1} describes the alignment *in time* of the vocal fold contacting events with the end of the first oscillation of the slowest vocal tract resonance. This would be like pushing the child on a swing and giving the second push the first time they swing back to you.¹⁹ The second harmonic ($2f_o$) aligned with f_{R1} allows the slowest vocal tract resonance to oscillate two complete times before the vocal folds contact. In our swing analogy, we would let the child return to us a second time before the next push. Notably, the oscillation of the vocal tract air mass and the arc of the swinging child both diminish in amplitude as twice as much time passes before they are once again set in motion. Donald Miller's second formant tuning strategy has no correlate in this analogy (the swing only has one dominant resonance), but the vocal tract tuned

for this resonance strategy simply leverages the oscillation of the next fastest resonance while quickly damping the slowest resonance (f_{R1}). In this way, the vocal folds oscillate at a *subharmonic* of the oscillating resonance.

This model explains two observable phenomena related to formant tuning. First, there appear to be no consequential effects of formant tuning when singing low pitches.²⁰ That is to say, although a spectrum may display the alignment of f_{R1} and $8f_o$ at a low pitch (for example), the amplitude of the pressure oscillations (resonances) in the vocal tract will have dropped dramatically, to the point that the effect on the next supraglottal impulse is questionable. The response of the vocal tract has died down because the pitch period is so long. Second, successful formant tuning likely does not solely impact the harmonic frequency in question. The ongoing resonance in the vocal tract as the folds contact likely affects the speed and strength of the supraglottal impulse generated with each vocal fold contacting event. This impacts the entire source impulse in a nonlinear fashion, contributing energy across the entire spectrum. For example, the strong second harmonic ($2f_o$) commonly observed in many approaches to belting²¹ likely points to a physical oscillation in the supraglottal air mass that may help to generate much faster aspects of the spectrum.

Why must formant tuning occur at whole-number multiples of the frequency of vocal fold oscillation? Because those are the resonant frequencies that best align drops in the pressure of the resonance with the subsequent supraglottal impulses. If the hypothetical frequency of our slowest resonance (f_{R1}) is three times faster than the frequency of vocal fold oscillation ($3f_o$), every third oscillation of the resonance will assist the periodic drop in supraglottal pressure. This will be true even if that resonance is a part of the complex oscillating pattern within the vocal tract that also features other rates of pressure oscillation. If our f_{R1} is $2.5f_o$ —imagine the second and third harmonics straddling the resonance in a spectrum—phonation continues reasonably unaffected. The resonance is excited at the start of each period, but its misalignment with

each subsequent glottal closing diminishes that energy by the end of each period. Three possible outcomes emerge: (1) The result will be less-efficient phonation as the supraglottal air mass no longer receives the acoustic boost of a resonance that aligns with each glottal contacting event; (2) the overall amplitude of the signal may decrease; or (3) the timbre may neutralize as the strong contribution of the tone colors of energy at $2f_o$ or $3f_o$ is reduced.²² Returning to the model of pushing a child on a swing, we can slightly misalign our pushing effort without completely arresting (damping) the swinging motion. Conceptually, strategies for f_{R2} tuning account for this kind of misalignment of f_{R1} by tuning the drop in pressure of an even faster resonance to another multiple of the frequency of supraglottal impulses.²³

Efficient phonation seems to be agnostic to which resonance best aligns with the glottal impulse, so long as sufficient acoustic energy exists in the vocal tract. Imagine this phenomenon as a resonance that oscillates over time rather than as the amplification of a persistent harmonic. This allows one to appreciate why formant tuning appears important at times but irrelevant at other times.

THE PITCH EXCEPTION OF OVERTONE SINGING

Pitch perception is delightfully complex. The remainder of this chapter will use a time-and-pressure-based model to discuss notable exceptions to the framework that was covered in [chapter 5](#). I argue here that some observable phenomena related to pitch are best conceptualized through their time and pressure characteristics.

Let us begin with *overtone singing*. Here this term points to the generation of a pure overtone in the context of otherwise periodic phonation. Other, often noisier approaches to overtone singing exist that leverage the same basic premise. We know that in periodic phonation in a speech-pitch range (see figure 6.5), we tend to see a drop in pressure at what we define as the beginning of each pitch period, and a drop in the amplitude of the waveform throughout the period. Imagine again giving a child a single push on a swing (and, to be sure, this waveform traces a complex swinging motion). The amplitude of their oscillation decreases in distance from the midline over time.

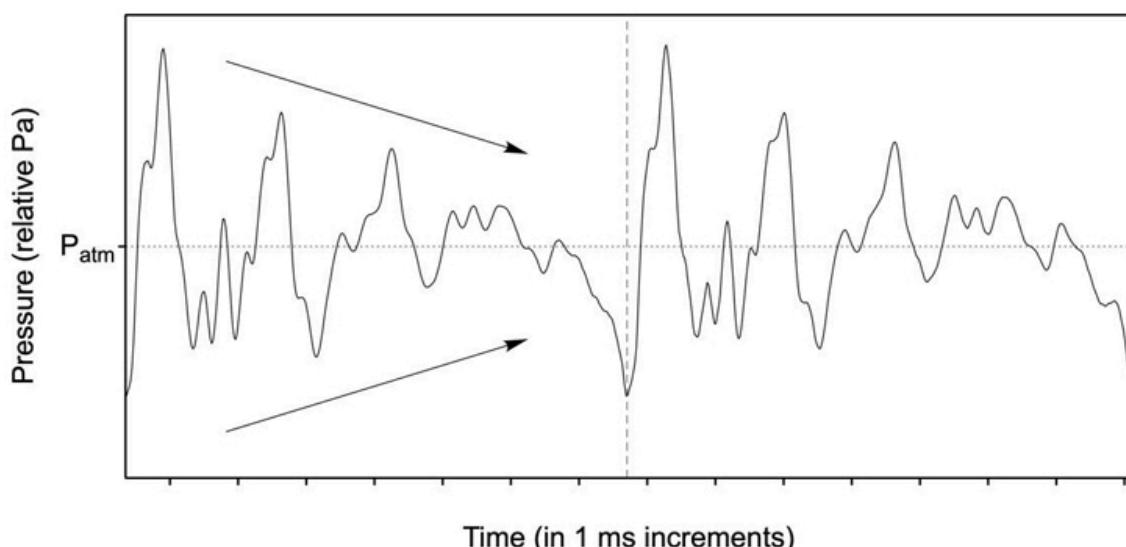


Figure 6.5 Two periods of a waveform of cisgender man singing [a] at pitch D_b3 (~135 Hz, 7.4 ms). Approximate start of second period marked with a dashed gray line. Arrows show the decay of pressure oscillations through the period as the pattern settles closer to the

midline (atmospheric pressure). Source: Recorded by author under controlled conditions.

Thinking back to the physiological process of auditory transduction covered in [chapter 3](#), imagine the pressure wave in the air setting in motion the tympanic membrane, the ossicles, and ultimately the final ossicle: the pistonlike *stapes*. The stapes instigates the wave that ripples through the fluid-filled cochlea, which sets in motion the frequency-filtering process of the inner ear. The start of each period introduces a strong pull/push force on the stapes which in turn transfers that motion to the cochlea. This may bring some nuance to what we see in an audio waveform: Peaks (excursions above the midline) represent “pushes” on the tympanic membrane and stapes, while troughs (excursions below the midline) represent “pulls.” Subsequent pulls and pushes within each period diminish in amplitude. This sets up a regular pattern of transients per period in [figure 6.5](#) that comports with pitch perception according to the timing theory.

Compare this with [figure 6.6](#), which shows a cisgender woman executing a D \flat 4 with a pronounced overtone at 5f $_o$. This pressure wave sets up a slow repeating pattern (once per period, marked with a dashed gray line) of transients giving rise to the pitch percept of D \flat 4. Note the faster ripple within those periods. Straightforward math tells us that this faster pattern has a frequency five times that of the slower pattern. Note too that the amplitude of this faster oscillation does not significantly decrease over the entire period. This is certainly unlike what we observe in figure 6.5.

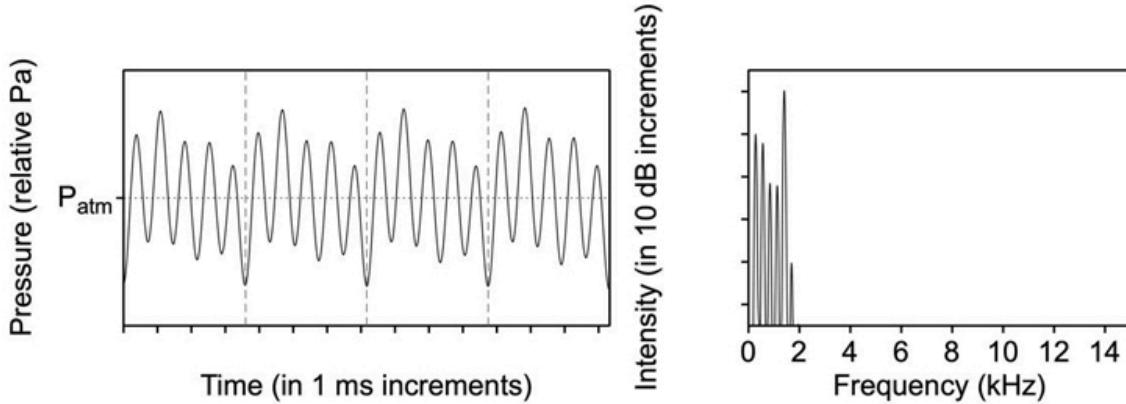


Figure 6.6 Waveform of a cisgender woman singing the pitch D b 4 with a pronounced overtone at $5f_0$. Approximate start of each period marked with a dashed gray line. Note that the overall deviations in amplitude from the midline remain high throughout the entire pitch period of D b 4. Source: Recorded by author under controlled conditions.

Consider for a moment what this pressure wave looks like as the *stapes* pushes into the fluid of the cochlea. Its complex, pistonlike motion traces two patterns at once (see figure 6.7). The gray waveform outlines the slower pattern, corresponding to the first and second harmonic ($1f_0$ and $2f_0$) in the spectrum. That pattern repeats once every 3.6 ms (D_b4, ~ 278 Hz). The black waveform outlines the continuous fifth harmonic ($5f_0$) in the spectrum, a sinusoidal motion repeating every 0.7 ms (~ 1.39 kHz).

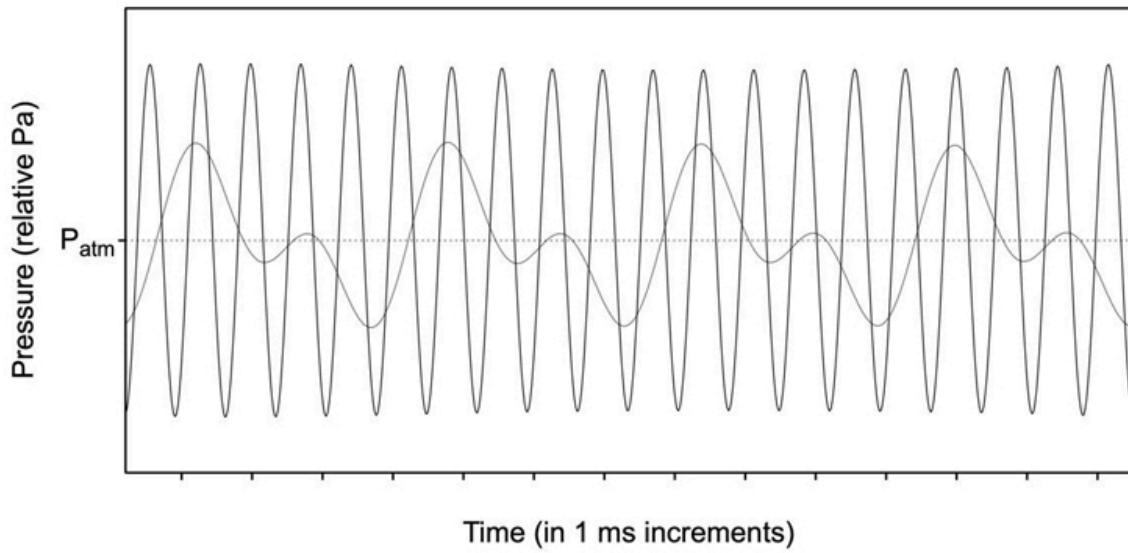


Figure 6.7 Overlay of 1fo and 2f_o (gray) and 5fo (black) from Figure 6.6. Source: Synthesized by author in Praat.

Now imagine the motion of the *stapes* as it traces this pattern forward and backward. The faster pattern regularly pushes and pulls on the stapes while that bone simultaneously moves in and out at the speed of the slower pattern. Add the amplitudes of the two waveforms in figure 6.7 together (allow them to constructively interfere), and the result is an approximation of the image shown in figure 6.6. Despite moving in and out slowly, the stapes also traces out the faster pattern *as though it were an additional fundamental*.

This leads to an intriguing explanation of overtone singing. The successful vocal tract posture for overtone singing needs to tune a resonance to an integer multiple of the pitch frequency; but it must also be shaped in a way that prevents that energy from significantly dropping in amplitude over the course of the period. This is achieved by decreasing the damping of the vocal tract for the resonance in question. In doing so, the oscillation continues to the end of the period. In a spectrum, this would appear as a strong harmonic. In my experience, singers tend to achieve this resonance adjustment at least in part by firming the tissue of the tongue. If this is so, it suggests the corollary that non-overtone

singing produces a more contiguous and gradual spectrum when *no single resonance completely dominates the entire pitch period*.

Strictly speaking, overtone singers do not amplify a higher harmonic of the fundamental. Instead, they excite a resonance of the vocal tract such that it (1) receives a well-timed boost by each supraglottal impulse, and (2) never significantly decreases in amplitude (dampens). The only frequencies that meet both of these physical requirements are integer multiples (harmonic frequencies) of the fundamental frequency. Therefore, strong overtones always fall into the harmonic series, even though they are products of the vocal tract.

The result of this process can be understood in this way: The radiated pressure wave creates two separate, simultaneous pitch percepts in the mind of a listener. One is a complex wave at the fundamental frequency of supraglottal impulses; the other is a pure and continuously oscillating higher resonance at a harmonic frequency of that fundamental. If perception dictates what sounds the human voice is capable of making (recall our question at the end [chapter 3](#) that cut to the heart of what it is to train a singer), overtone singing is an example of a non-dysphonic voice producing two pitch percepts at once.

If you can overtone sing, explore this idea experientially in Lab #7.

THE PITCH EXCEPTION OF SUBHARMONIC SINGING

While overtone singing is an approach to phonation that nests a higher pitch percept (continuous oscillation of a resonance) within the longer period of a slower fundamental, *subharmonic* singing requires that the singer create a periodic sound with a much longer period than the vocal folds might otherwise normally sustain. I would like to point to two ways to conceive of subharmonic singing: (1) by examining the resulting pressure wave and how one perceives its pitch, and (2) by scrutinizing what is likely happening at the level of the glottis to create this pressure wave.

Figure 6.8 shows a waveform and spectrum for an [a] sung by a cisgender man on the pitches D_b3 (top, ~140 Hz) and D_b2 (bottom, ~70 Hz). D_b3 (top) is normal modal phonation. The D_b2 (bottom) is subharmonic phonation. As with figure 6.5, note in the sung D_b3 that the amplitude of the waveform is greatest at the start of each period and decreases over time. This is a common feature of modal singing, and a common pattern in the singing voices of endogenous testosterone puberty singers in this pitch range. Compare this with the waveform of subharmonic phonation in which there appears to be two excitation phases per period. The primary phase features an initial impulse that excites and dampens within the vocal tract as expected. The secondary phase features a small second boost that introduces new energy into the vocal tract. The pitch percept is generated by the full length of this pattern, which is about 14.2 ms in duration, or 70 Hz. But perhaps counterintuitively, the folds are not neatly opening and contacting once every one-seventieth of a second.

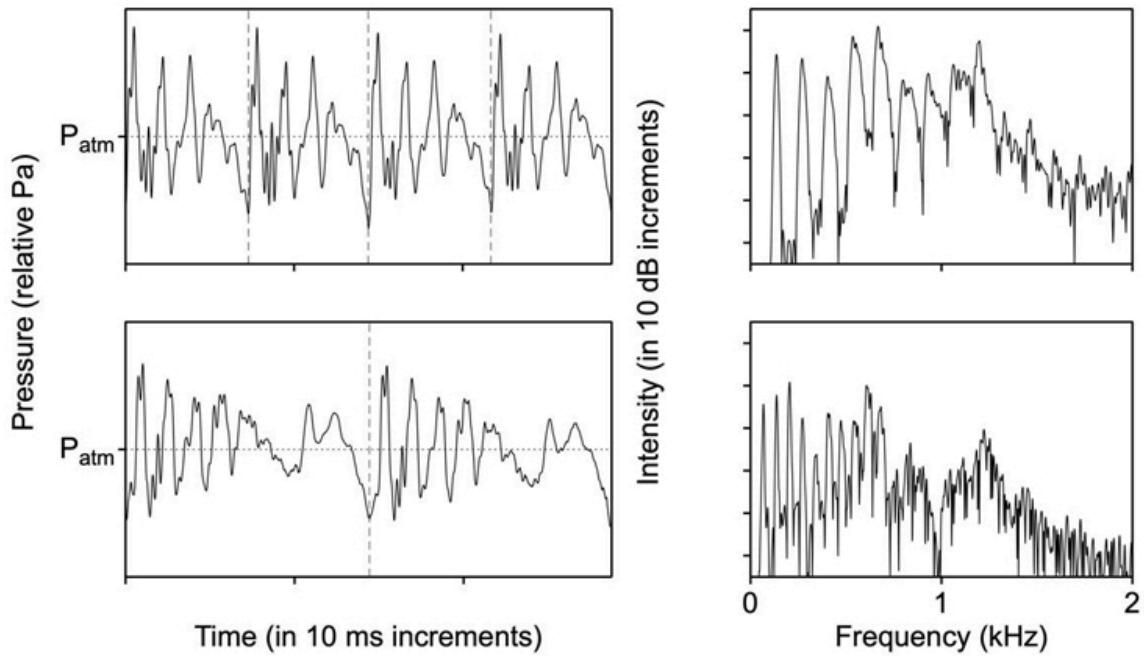


Figure 6.8 (Top left) waveform and (top right) spectrum of an [a] sung by a cisgender man in normal modal phonation at the pitch D₂₃ (~140 Hz, 7.1 ms). (Bottom left) waveform and (bottom right) spectrum of an [a] sung by a cisgender man in subharmonic phonation at the pitch D₂₂ (~70Hz, 14.2 ms). Approximate start of each period marked with a dashed gray line. Source: Recorded by author under controlled conditions.

As explored by Jan Švec et al. (1996), non-dysphonic vocal folds permit self-sustaining oscillation in patterns where the vocal folds oscillate with asymmetrical complexity within the glottal cycle. This complexity generates a transglottal airflow pattern that creates the strongest transient whenever the vocal folds meet at their midline.²⁴ When we align the waveforms of normal phonation and subharmonic phonation (see figure 6.8, top and bottom), their periods match exactly in a 1:2 ratio (meaning the pitch percept of subharmonic phonation is an octave below the modal singing pitch), as predicted by Švec et al.

Subharmonic phonation appears to leverage the timing theory of pitch perception. Because the transient at the rate of the longer period is stronger than any subsequent within-period change in pressure, the

percept becomes the pitch of the longest period, although the vocal folds are indeed oscillating faster.

Any non-dysphonic singer can likely do this, regardless of hormonal regime or age. Infants and young children who lack the vocal motor control of adults may easily slip into subharmonic phonation. Often mistaken for vocal fry, subharmonic phonation routinely punctuates our speech as well. Once you learn to see the telltale *additional harmonics* of a pitch period an octave lower on a spectrogram or hear the sudden drop of the pitch by an octave, you will be able to identify it with ease (see figure 6.9). Additional subharmonics more than an octave below the intended pitch are also possible, warranting further research.

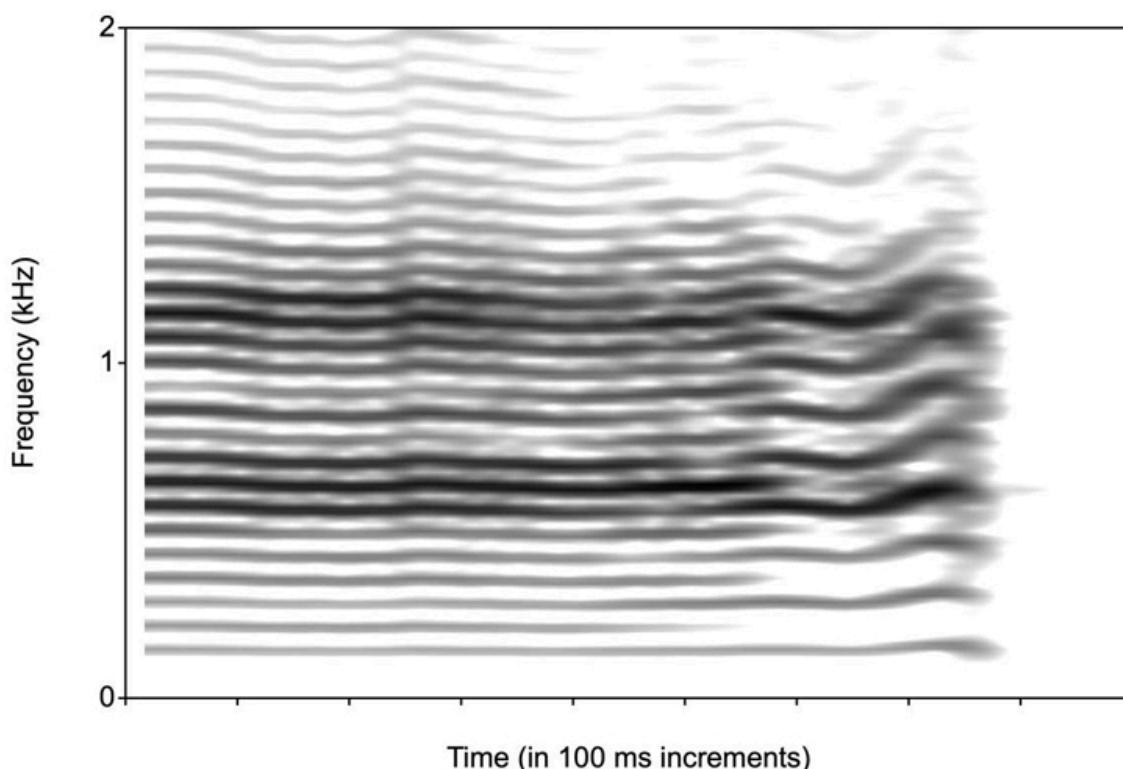


Figure 6.9 A longer time view of the sample that generated Figure 6.8. Subharmonic phonation followed by a switch to modal phonation. Note that subharmonic phonation has additional harmonics between the harmonics of the modal phonation.

THE TIMBRE EXCEPTION OF VOCAL FRY

So far, we have investigated normal voicing, overtone singing, and subharmonic phonation. While these are by no means the only ways to create a sung sound with a pitch percept, these contrasting modes of phonation do demonstrate that the human voice can produce pitch percepts in multiple ways. A fourth phonation type, *vocal fry*, throws into relief exactly how important periodicity is to pitch, and how timbre and pitch may be separated.

Vocal fry is broadly considered to be the “lowest” laryngeal vibratory mechanism of the voice.²⁵ It is a creaky, popping, pulsing, pitch-less sound.²⁶ Notably though, vocal fry has timbre. Indeed, Donald Miller et al. (2000) suggest utilizing vocal fry as a nonperiodic method for extracting accurate vocal tract formant frequency measures.²⁷ By definition, those formants help to characterize timbre.

Consider the following images ([figure 6.10](#)), illustrating vocal fry of an [a] by the same singer in figure 6.8. According to the transient theory of voice production, the dampened resonant response of the vocal tract to each impulse should be reasonably similar in figure 6.10 as it was in both the modal and subharmonic singing in figure 6.8, since the vocal tract is shaped similarly in all three cases. Figure 6.10 (top left) shows a zoomed-in section of the vocal fry waveform. Note that these ripples indeed follow a very similar pattern to those found in both modal and subharmonic phonation in figure 6.8. This is because the “pops” of vocal fry *also* excite the resonant response of the vocal tract with transient changes in pressure.

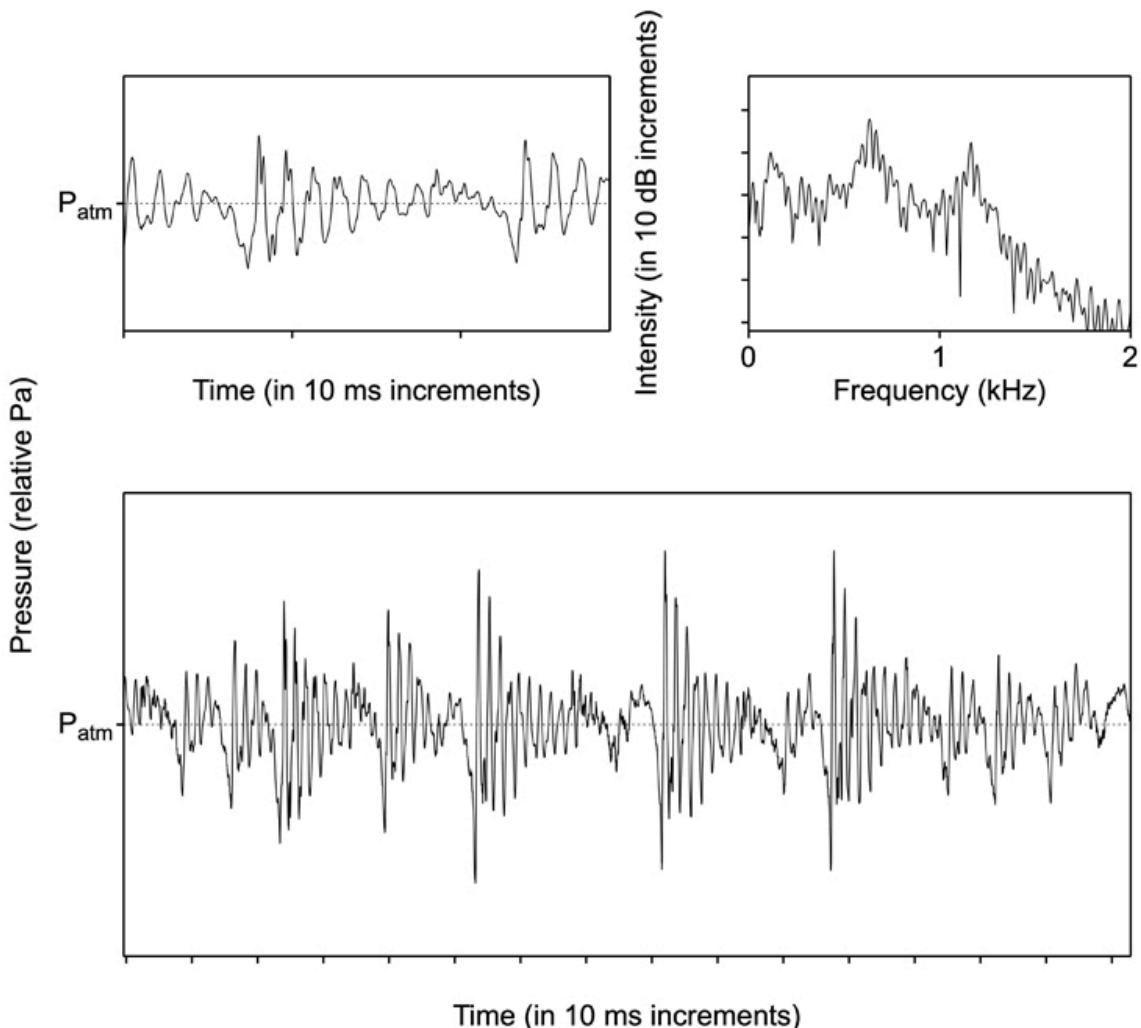


Figure 6.10 Vocal fry of an [a] by the same singer as in Figure 6.8. Top left is zoomed into show several excitations of the vocal tract resonances on the same timescale of Figure 24. Top right shows a spectrum, and bottom zooms out to show several aperiodic excitation patterns. Source: Recorded by author under controlled conditions.

Note, though, that when zoomed out (figure 6.10, bottom), there is no consistent duration between each transient. The resonances play out similarly after every transient, but those transients are aperiodic. This is also visible in the spectrum (figure 6.10, top right), which shows peaks at ~ 650 Hz and $\sim 1,150$ Hz, as were present in both modal and subharmonic phonation in figure 6.8. This is how vocal fry conveys the timbre identity of the vowel without pitch. It is aperiodic excitations of a vocal tract that

could be excited periodically. In a way of thinking, vocal fry more completely transmits the timbral potential of all the vocal tract resonances, because the randomized duration between glottal impulses at times allows for a much longer oscillation of the resonances within the vocal tract than found in periodic phonation.

If you would like to explore this idea experientially, carry out Lab #8.

THE PITCH EXCEPTION OF NOISY DISTORTION

Some distorted or extreme sounds allow us to tie together several of the above concepts in a novel way. While distorted means of phonation have most recently been associated with rock and heavy metal singing, Mauro Barro Fiúza et al. (2018) suggests that these “intentional vocal distortions” have long been employed in singing across disparate cultures.²⁸ Melissa Cross (2019) suggests that fry screaming involves the aperiodic “fluttering” of the true vocal folds.²⁹ Daniel Zanger Borch et al. (2003) suggest that distorted singing more broadly features “vibrations of the supraglottal mucosa.”³⁰ Regardless of the manner in which the vocal tract is excited in extreme sounds, I would like to explore how a pitch-like percept can manifest in the absence of any periodic oscillation of the vocal folds.

Let us return for a moment to the James Brown sample from [chapter 5](#) ([figure 6.11](#), left). Note that the waveform of the /æ/ is characterized by a quasi-regular pattern. Though it is clearly random and aperiodic, there is still a ripple that oscillates somewhat regularly. This happens around 1,100 Hz, and it becomes a dominant frequency in the otherwise noisy percept. Zooming out to a slightly longer timescale (figure 6.11, right) shows us that this basic pattern persists, although there are neither periodic repetitions nor a regular pattern of damping. It is in this way that the sound he makes generates a noisy, quasi-pitch percept by continually energizing a vocal tract resonance with noise in the absence of a periodic glottal impulse.

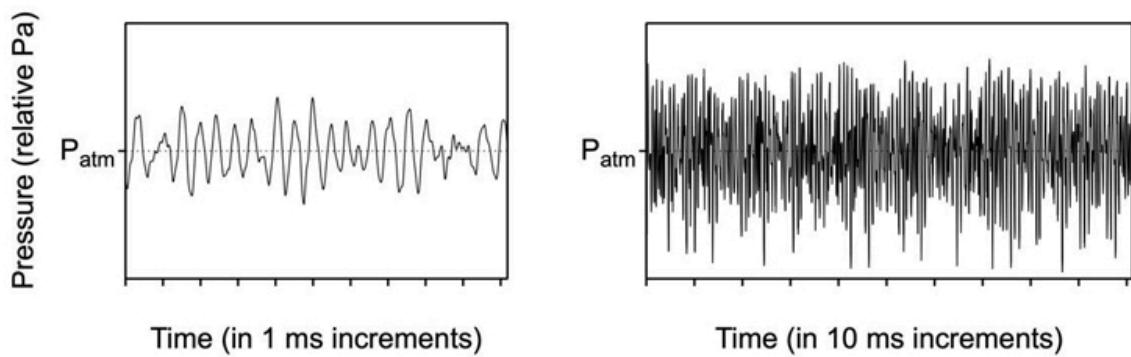


Figure 6.11 Waveform of James Brown fry scream [wæu] from “I Got You (I Feel Good)” from Figure 5.10. Zoomed in (left) and zoomed out (right). Source: https://www.youtube.com/watch?v=W-rn7i_ETYc.

CONCLUSIONS AND POTENTIAL APPLICATIONS

By centering a time-and-pressure domain view of voice production, I hope to offer a complementary model to understand real, observable phenomena. At times, exploring this model leads to uncomfortable suggestions: notably that harmonics do not appear to temporally precede resonances as laid out in the simplest version of the source-filter model. Please keep in mind that we separate the contribution of the vocal folds from the contribution of the vocal tract for convenience at a model level, not for accuracy. When you encounter images showing the spectrum of the vocal fold source sound with discrete harmonics of decreasing intensity that are ready to be reshaped by the vocal tract transfer function, please know that this discretization does not exist in nature. It is a useful way to conceive of the separate processes of voicing. On the other hand, the resonances of the vocal tract do in fact seem to be excited as real physical oscillations of the air mass in the vocal tract per glottal cycle. The repetition of that resonant response *is* the periodic pattern that can be analyzed as a harmonic spectrum. As established above, this suggests that *timbre has duration*. To grasp this, the time domain must be confronted.

It is worth reflecting on why spectrum- and spectrogram-based models continue to dominate the teaching of voice acoustics when they potentially pull us away from other, complementary models. By no means am I suggesting that everyone who teaches a voice acoustics unit in a voice pedagogy or vocology class is confused about this. But it is my opinion that such dominant models are the most pedagogically helpful when we remember that they conveniently display repeated aspects of the information that is also available via time-and-pressure-based models.

For example, if one observes that there is a strong third harmonic ($3f_0$) in a singer's sound, a spectrum will display its frequency far more intuitively than a waveform. Where we may lose our footing is in forgetting that that spectrum shows repeating aspects within the pitch

period of the complex wave in terms of harmonic frequencies of the slowest repeating pattern—in other words, the fundamental. The acoustic energy the spectrum represents as that third harmonic exists on the timescale of a single period as a resonance that ripples at one-third the duration of the pitch period (or three times the frequency in Hz of the fundamental). That is to say, there will be three noticeable ripples per pitch period in the waveform. That resonance will generally decrease in amplitude through the duration of the period, which directly impacts the resulting percept.

This nuance is typically summarized away in a spectrum or spectrogram because the algorithm that generates that image typically averages information over a window of time that is much longer than the pitch period. The $3f_0$ in the spectrum does not point to a pure tone that independently exists in time and space. The energy of that harmonic likely does not even consistently oscillate for the entire pitch period (overtone singing being a notable exception).

The timing theory of pitch perception makes more sense once we consider the repetition of transients in a waveform, as opposed to the fundamental in a spectrum. The waveform of a sound with a weak or missing fundamental ($1f_0$) in that spectrum still repeats at the same period duration as that same sound would have with a strong fundamental. The interdependent nature of pitch and timbre similarly becomes obvious when we see them interact in a waveform.

Many questions and answers arise from this discussion: Why can you not separately hear all of those harmonics on the spectrogram? What is the nature of the vocal fold source sound? What are resonances, and why can you move them or align them with harmonics? Why does timbre change with pitch? What are harmonics, and where do they come from? What process creates the separately heard pure tone when pass-filtering just one harmonic?

I believe that much of what remains confusing about voice acoustics begins to clear up when viewed through the time-and-pressure domain. At a minimum, we should consider what ideas may flow from all available models.

Why include this view of voice acoustics in a book about perception? As we discovered in [chapter 3](#), the physical limitations of the process of auditory transduction are a fruitful launch pad from which to understand voice perception. The sooner one can develop an intuition for how the acoustic pressure pattern is created, the sooner one can understand how different functional adjustments trigger specific perceptual qualities.

DISCUSSION QUESTIONS

- Why does modern voice pedagogy rely so heavily on the steady-state theory of voice production?
- What mathematical process generates the output of a spectrogram (the image that the software creates)?
- In a spectrum or spectrogram, we are accustomed to seeing the effect of resonances on harmonics. But at what timescale and in what manner do we see resonances in a waveform?
- Given the idea of wave interference, under which conditions would the resonance information of a single period of voicing influence the waveform of the following period of voicing?
- Does timbre have duration? Support your answer.
- Discuss pitch, resonance, and timbre in the context of the examples offered: formant tuning, overtone singing, subharmonic singing, vocal fry, and fry scream.

NOTES

1. Sean A. Fulop, *Speech Spectrum Analysis* (Berlin: Springer, 2011).
2. "Transient," *The Oxford English Dictionary*, <https://www.oed.com/search/dictionary/?scope=Entries&q=transient>.
3. "Transient," *Oxford Reference*, <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803105341303>.
4. "Steady-state," *Oxford Reference*, <https://www.oxfordreference.com/display/10.1093/acref/9780198832102.001.0001/acref-9780198832102-e-6142?rskey=2Ylo4v&result=4>.
5. Harvey Fletcher, *Speech and Hearing* (New York: D. Van Nostrand, 1929), 49–50.
6. Ingo R. Titze, *Principles of Voice Production* (Iowa City: National Center for Voice and Speech, 2000), 99–113.
7. Gordon E. Peterson and Harold L. Barney, "Control Methods Used in a Study of Vowels," in *The Journal of the Acoustical Society of America* 24, no. 2 (March 1952): 177.
8. Stephen F. Austin, "Canaries in the Coal Mine: The Pure Vowel," in *Journal of Singing* 68, no. 1 (September/October 2011): 83–85; Stephen F. Austin, "Carlo Bassini's The Art of Singing, Part 1," in *Journal of Singing* 66, no. 5 (May/June 2010): 591–98.
9. Kenneth Bozeman, *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013), 34, 85.
10. Donald G. Miller, *Resonance in Singing: Voice Building through Acoustic Feedback* (Princeton, NJ: Inside View Press, 2008), 96–105.
11. Berton Coffin, *Coffin's Sounds of Singing: Principles and Applications of Vocal Techniques with Chromatic Vowel Chart*, 2nd ed. (Lanham, MD: Scarecrow Press, 2002).
12. Bozeman, *Practical Vocal Acoustics*.
13. Titze, *Principles*, 257–58, 303.
14. Ingo R. Titze et al., "Toward a Consensus on Symbolic Notation of Harmonics, Resonances, and Formants in Vocalization," in *The Journal of the Acoustical Society* 137, no. 5 (May 2015): 3005–7.
15. Bozeman, *Practical*, 23–24.
16. Donald Miller, *Resonance in Singing*, 52–53.
17. Bozeman, *Practical*, 68–70, 115; Miller, *Resonance*, 54–55.
18. Miller, *Resonance*, 1–5, 66, 71, 75–77, 88, 111.
19. Titze, *Principles*, 91–92.
20. Daniel A. H. Mitton, "Sung Russian for the Low Male Voice Classical Singer: The Latent Pedagogical Value of Sung Russian" (DMA diss., University of

Toronto, 2020), 131.

21. Miller, *Resonance*, 88.
22. Bozeman, 26–27. This is likely a good explanation of the closing events characterized by Bozeman as the *pitch of turning*.
23. Miller, *Resonance*, 88.
24. Jan G. Švec, Harm K. Schutte, and Donald G. Miller, "A Subharmonic Vibratory Pattern in Normal Vocal Folds," in *Journal of Speech and Hearing Research* 39, no. 1 (1996): 135–43. DOI: 10.1044/jshr.3901.135.
25. Bernard Roubeau, Nathalie Henrich, and Michèle Castellengo, "Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited," in *Journal of Voice* 23, no. 4 (2007): 425–38.
26. Titze, *Principles*, 283–84. Any pitched sound $f_0 < \sim 70$ Hz will exhibit a pulse-like quality despite being periodic. The phenomenon explored here is the aperiodic version.
27. Donald G. Miller, Arend M. Sulter, Harm K. Schutte and Rienhard F. Wolf, "Comparison of Vocal Tract Formants in Singing and Nonperiodic Phonation," in "Registers in Singing: Empirical and Systematic Studies in the Theory of the Singing Voice" (PhD diss., University of Groningen, 2000), 188.
28. Mauro Barro Fiúza and Marta Assumpção de Andrada e Silva, "Can Singing with Rasp Be a Healthy Practice?" *Distúrbios da Comunicação* 30, no. 4 (December 2018): 802–808, <http://dx.doi.org/10.23925/2176-2724.2018v30i4p802-808>.
29. Melissa Cross, "The Zen Of Screaming 2 Trailer!! New DVD!" [3:31], YouTube, <https://www.youtube.com/watch?v=spZWQwxNKHg>
30. Daniel Zanger Borch, Johan Sundberg, P. A. Lindestad, M. Thalén, "Vocal Fold Vibration and Voice Source Aperiodicity in Phonatorily Distorted Singing," in *Department for Speech, Music and Hearing Quarterly Progress and Status Report*, KTH, 2003, 89.

7

Perceptual Qualities of Auditory Roughness and Pitch Resolution*

The preceding material argues that different portions of a spectrum (the slower and faster oscillations within each pitch period) trigger different perceptual qualities. If this is true, we may explore what those qualities are, label them, *and attempt to relate them to voice function* by asking which aspects of voice production produce those slower and faster oscillations.

To summarize from [chapter 5](#), the portion of the spectrum unresolved from the pitch is not just noise; it is experientially more *buzzy*. The resolved portion is not just pitch; it is experientially *more pure*. However, while pitch resolution dissipates above approximately the eighth harmonic, the transition from pure to buzzy in otherwise periodic phonation typically takes place lower in the harmonic series: around the fifth to sixth harmonic for much of the singable range.^{1 , 2 , 3 , 4} This buzzing quality is termed “auditory roughness.” This chapter will address this perceptual quality and its physiological cause.

WHAT IS AUDITORY ROUGHNESS?

There is the potential for confusion whenever two fields of study use a similar term to describe different phenomena. A wide range of sounds may be described as *rough*, many of which likely leverage similar innate reactions of our hearing mechanism. Indeed, in speech pathology, the term “roughness” labels an undesirable, uneven sound caused by aperiodic noise in the voice signal.⁵ We frequently attempt to rehabilitate this when it interferes with our ability to sing or communicate. Within artistic singing practices, “rough” can also point to technical approaches that allow one to sustainably produce extreme sounds like grunts, growls, and screams.⁶ And many singing artists split the difference, forming their characteristic sound around a rough quality that another singer might want to reform.

In the study of psychoacoustics, the compound term “auditory roughness” may be used interchangeably with “roughness,” but in both cases these terms point to a perceptual phenomenon. In this book, both terms point to the physiological and perceptual process that gives rise to these rough perceptual qualities rather than an overall voice quality *per se*. Of course, since the ear helps create the sound of a singer, it is difficult to parse out these distinctions. The buzzy perceptual quality of auditory roughness explored here is distinct from the roughness one might associate with myriad extreme artistic choices, pathologies, or even functional deficits. It is, however, present as a normal, familiar, desirable buzzing quality that we tend to cultivate in robust singing.⁷

The reader will remember from [chapter 3](#) that the basilar membrane has frequency resolution limitations. Moving from the physiological to the perceptual, as counterintuitive as this might be, the percept associated with multiple stimuli falling into a critical band is a *buzzy* sound. Next, I will describe how to explore and experience this percept in isolation.

Different authors explain auditory roughness differently. Typically, they frame the phenomenon in terms of the distance between two simultaneously presented pure stimuli. This distance is characterized in

either Hz or musical intervals. In a physical sense, auditory roughness is a phenomenon that exists because of the physical proximity of the separate basilar membrane responses to those separate stimuli. In practical terms, we may both calculate the width of the proximity of this effect for specific lengths of the basilar membrane in Hz, or we may think more simply, using musical intervals. Both measurements ultimately point to a physical distance along the basilar membrane.

Howard and Angus (2017) frame auditory roughness in reference to the beating quality created when two equally loud pure tones are less than 12.5 Hz apart. As the two tones approach a unison, the beating slows until it disappears. When separated by greater than around 15 Hz, this beating quality goes away and becomes a fused percept with a rough quality. As the distance in frequency between these two tones continues to grow, the rough percept disappears, replaced by the percept of two clearly separate tones.⁸

Heller (2013) uses the umbrella terms *self-dissonance* and *autodissonance* to capture both beating and roughness, and he places this in a musical context.⁹ If a musical interval would be close enough to trigger a dissonant musical sound, the higher harmonics of a complex tone with the same intervallic relationship to one another will similarly sound “self-dissonant.” He points to intervals closer than a perfect fourth as the threshold for this phenomenon. He also cautions that we cannot assume that the higher harmonics of a periodic sound will behave as independent pure tones do, suggesting that “a perfectly fine, rich tone may be dissected for dissonant partners contained in the sound, yet the dissonances are not evident in the whole tone.”¹⁰ Similarly, Sundberg (1987) frames “roughness of the timbre” in a voice in terms of the position of harmonics within the harmonic series, suggesting that auditory roughness can only arise from “pairs of partials above the fourth partial.”¹¹ The fifth and sixth harmonics are less than a major third apart, with every subsequent pair of harmonics separated by smaller and smaller musical intervals. Thus, only the higher frequency (faster oscillations) harmonics of a periodic sound trigger auditory roughness.

Again, we find two different framings of auditory roughness: in relation to distance between tones in Hz and distance between tones in musical intervals. Perhaps we can note that the term *beats*, made distinct in Howard and Angus, describes an additional rough phenomenon characterized by wave-interference pulses that fall below the lower threshold of pitch perception (under about 20 Hz). As I will explore below, such pulses are rarely a feature within a single singing voice. Thus, a musical interval between adjacent harmonics may make more sense as a rule of thumb than a frequency-difference threshold.

Here are a few examples: Consider that two separately presented tones at 100 Hz and 106 Hz (approximately one half step apart) would produce wave-interference beats six times per second (see figure 7.1, top) and would sound rough. In fact, two tones at 200 Hz and 206 Hz will also beat six times per second and sound rough, although it turns out that their musical interval is narrower than a half step (see figure 7.1, middle). This is likely counterintuitive, but remember that pitch perception is nonlinear. Doubling or halving the frequency raises or lowers the pitch by an octave. Adding around 6 percent to the frequency raises the pitch a half step. The frequency difference between the two pitches in a musical interval will always depend on the starting frequency. Musical intervals are ratios, not fixed changes in frequency.

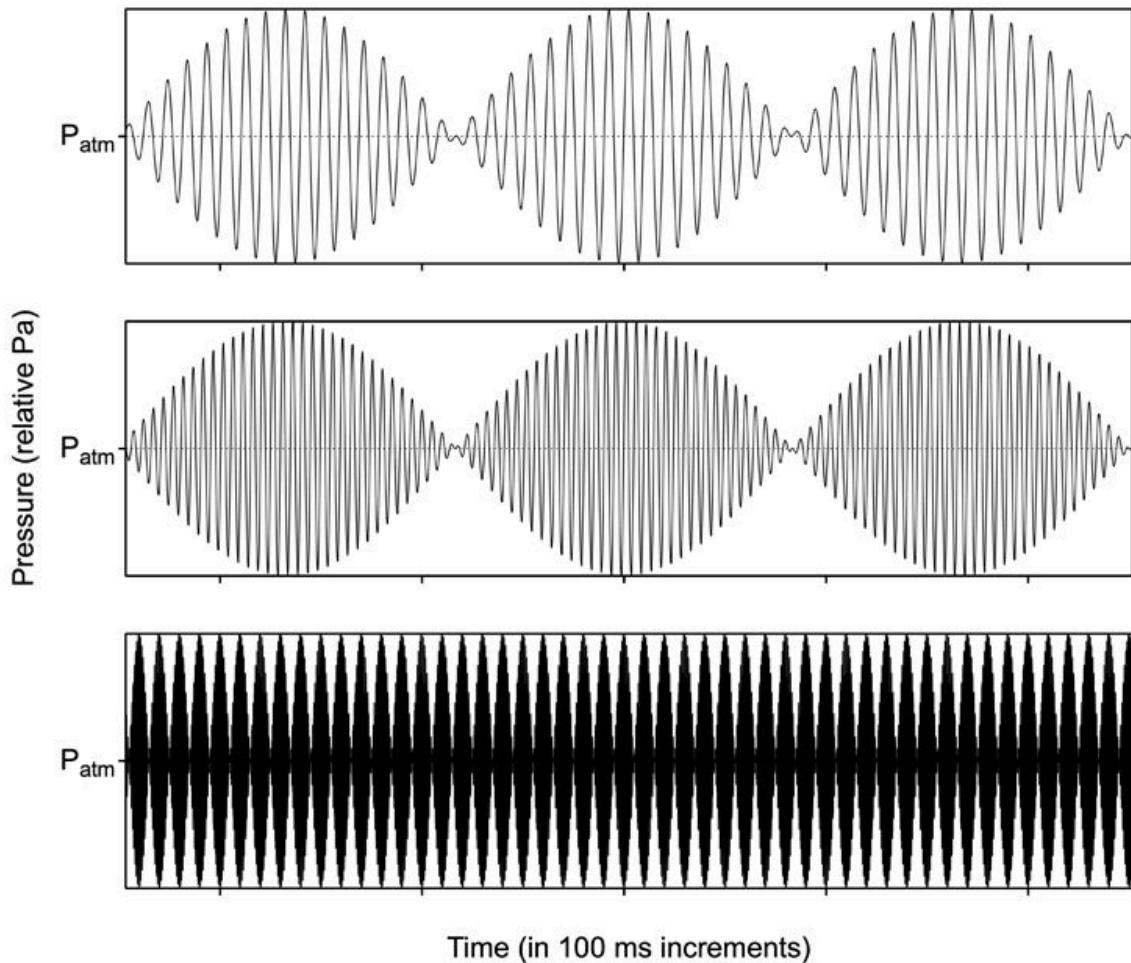


Figure 7.1 (Top) 6 Hz pulsing pattern formed by simultaneous pure tones at 100 Hz and 106 Hz. **(Middle)** 6 Hz pulsing pattern formed by simultaneous pure tones at 200 Hz and 206 Hz. **(Bottom)** 100 Hz pulsing pattern formed by simultaneous pure tones at 1.6 kHz and 1.7 kHz. Source: Synthesized by author in Praat.

Continuing this exploration, the sixteenth and seventeenth harmonics of a tone at 100 Hz (1,600 Hz and 1,700 Hz, respectively) are *also* approximately a half step apart, but they do not trigger the same beats as the previous two examples did (see [figure 7.1](#), bottom). This is because their interference pattern pulses at 100 Hz, which falls well within the range of human pitch perception. We hear that pulsing as a pitch. In fact, for adjacent harmonics of a periodic tone to produce beats, the fundamental would have to be less than 15 Hz. This would sound like a

series of regularly repeating pulses rather than a pitch. So in a voice, Heller's autodissonance always produces pulses; we just hear pulses faster than about 20 Hz as a pitch.

As mentioned above, when two pure tones approach each other in frequency, strong pulses or beats are perceived at the frequency of the difference in their frequencies. Such beats have both a correlate in the physical world and a cognitive aspect. When combined in the air, two tones separated by 6 Hz (e.g., 100 Hz and 106 Hz) will form an interference pattern that pulses six times per second (see again figure 7.1, top). Since this pulsing is slower than 20 Hz, it will generate a rhythmic rather than a pitched percept.

We can also explore a less pronounced beating quality called *binaural beats* by playing these two separate tones over headphones, one in each ear. The emergence of binaural beats while listening to slightly different tones in each ear implies a cognitive mechanism is at work. We can be confident in this inference because the two pressure patterns never have the chance to interfere in the air.¹²

EXPLORING AUDITORY ROUGHNESS EXPERIMENTALLY

Auditory roughness may be easily explored by creating two pure tones in a tone-generating program or app of your choice, such as Audacity, Madde, VoceVista Video Pro, Praat, or any other. Hold the frequency of the first tone constant at C5 (~523 Hz) and slowly bring the second tone from G5 (~784 Hz) down to E_b5 (~622 Hz). As you bring the second tone closer in frequency, note that you will begin to perceive a dissonant buzzing. If the tones are sufficiently intense, you will be able to perceive the *difference tone*. The difference tone is the pitch equivalent to the difference between those frequencies. This suggests that a tone at 1,600 Hz and another at 1,700 Hz will generate a lower pitch percept at 100 Hz.

As two pure tones approach each other in frequency, and as that difference tone falls below the threshold on human hearing, they will begin to produce the above-mentioned *beating* sound like a metrical pulsing of intensity. In a low enough frequency range, this is strongly noticeable at intervals smaller than a half step. If the frequencies of the two tones were to double (to rise an octave in pitch), the musical interval would have to become larger in order to pulse at the same rate. An octave lower and that musical interval would have to be smaller. As the tones become even closer to one another, the beats slow down until they collapse into a perfect unison.¹³ Explore all these examples in Lab # 9.

Next we will explore how auditory roughness might exist in a truly complex sound. To do this we will generate a sawtooth wave at the pitch C4 (~262 Hz). VoceVista Video Pro, Praat, and Madde will all do this. First, generate that sawtooth wave and record it. Pass-filter the lowest five harmonics of the spectrum within one of these programs. This means you will *allow* those harmonics to pass while stopping the playback of higher harmonics. Note that this lower portion of the spectrum lacks roughness. Now pass-filter from the fifth harmonic ($5f_0$) upward. Notice how buzzy the quality is. If the pitch of your sawtooth falls within the pitch range in which humans tend to sing, the rough division of the spectrum at the fifth

harmonic will usually separate the pure sound from the rough sound.¹⁴ See Lab #10.

A loose definition of *auditory roughness*, then, suggests that in terms of musical intervals, any two separate pitched tones a minor third or closer will generally fall within a critical band of hearing. The rule of the minor third may be applied to harmonics of a voice as they appear on a spectrogram.¹⁵ Generally speaking, the percept of auditory roughness in a single voice will always be buzziness, never beats.

VISUALIZING CRITICAL BANDS OF HEARING

Let us return to the idea of the spatially organized frequency response of the basilar membrane within the cochlea (see “How Does the Ear Work?” in [chapter 3](#)). We have learned that different frequency pure tones will find resonance with different physical locations along this membrane on a continuum from high to low frequency. Closer in frequency is physically closer along the surface of the basilar membrane. But no matter how narrow the frequency band of the stimulus, the basilar membrane has a minimum displacement length. Even a pure tone displaces a short region of the basilar membrane, not just a single point. It is possible to simultaneously stimulate two adjacent portions of the basilar membrane, and the wash of high-frequency energy rippling through each pitch period creates overlapping areas of stimulation along the basilar membrane. This is the physical correlate to a critical band of hearing.

This is analogous to familiar resolution limitations of our eyes. We can see far enough that we can no longer make out the details of distant objects. The details are *there in nature*, but the resolution limitations of our eyes obscure that information. Similarly, it is possible to simultaneously stimulate two points along the basilar membrane that are so physically close that these stimuli overlap, even if these stimuli are not persistent at all timescales. The brain becomes confused as to whether two separate stimuli are present, and we experience it as a single, fused percept. With this understanding, we can acknowledge that the limitations of the human ear belie detail that may be present in nature.

In the harmonic model we can think about critical bands as if they are a line of bins (see [figure 7.2](#)). If each bin contains a single input, it passes that input along intact. If a bin has more than one input, it sums those inputs and passes through some amount of the buzzy quality associated with auditory roughness. In this model, think of the input arrows as separate harmonics in the harmonic series as they would appear in a spectrum or spectrogram. The output arrows show a single, increasingly buzzy percept per bin.

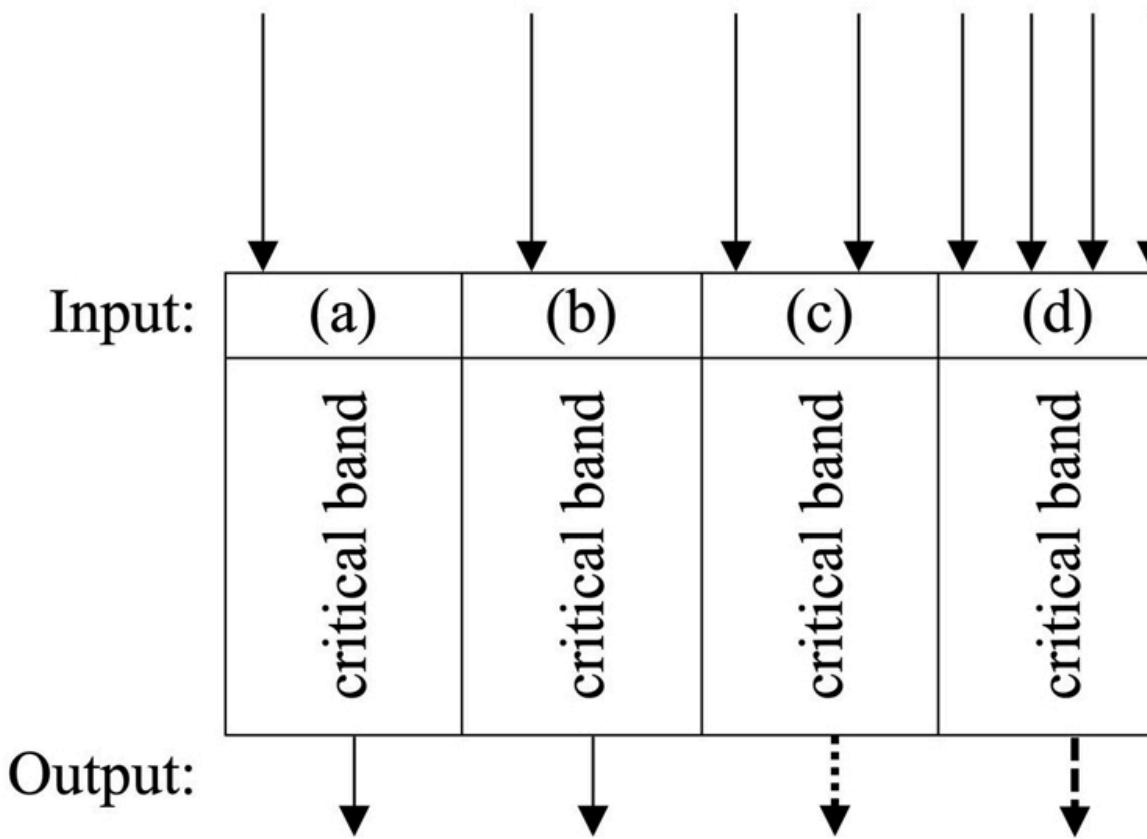


Figure 7.2 Author's schematic of four critical bands of hearing. Each critical band sums the input into a single output. Bands (a) and (b) exhibit no auditory roughness. Bands (c) and (d) exhibit increasing auditory roughness.

This suggests what initially may be an uncomfortable idea about singing voice perception: It is generally not possible to perceive individual harmonics above about the fifth harmonic. From a practical perspective, a spectrum or spectrogram would feel more accurate if it clearly showed the lowest five harmonics while blurring the higher ones (see figure 7.3).¹⁶ Interesting exceptions to this rule relate to the intensity of an individual harmonic relative to its surrounding harmonics, notably as found in some forms of overtone singing.

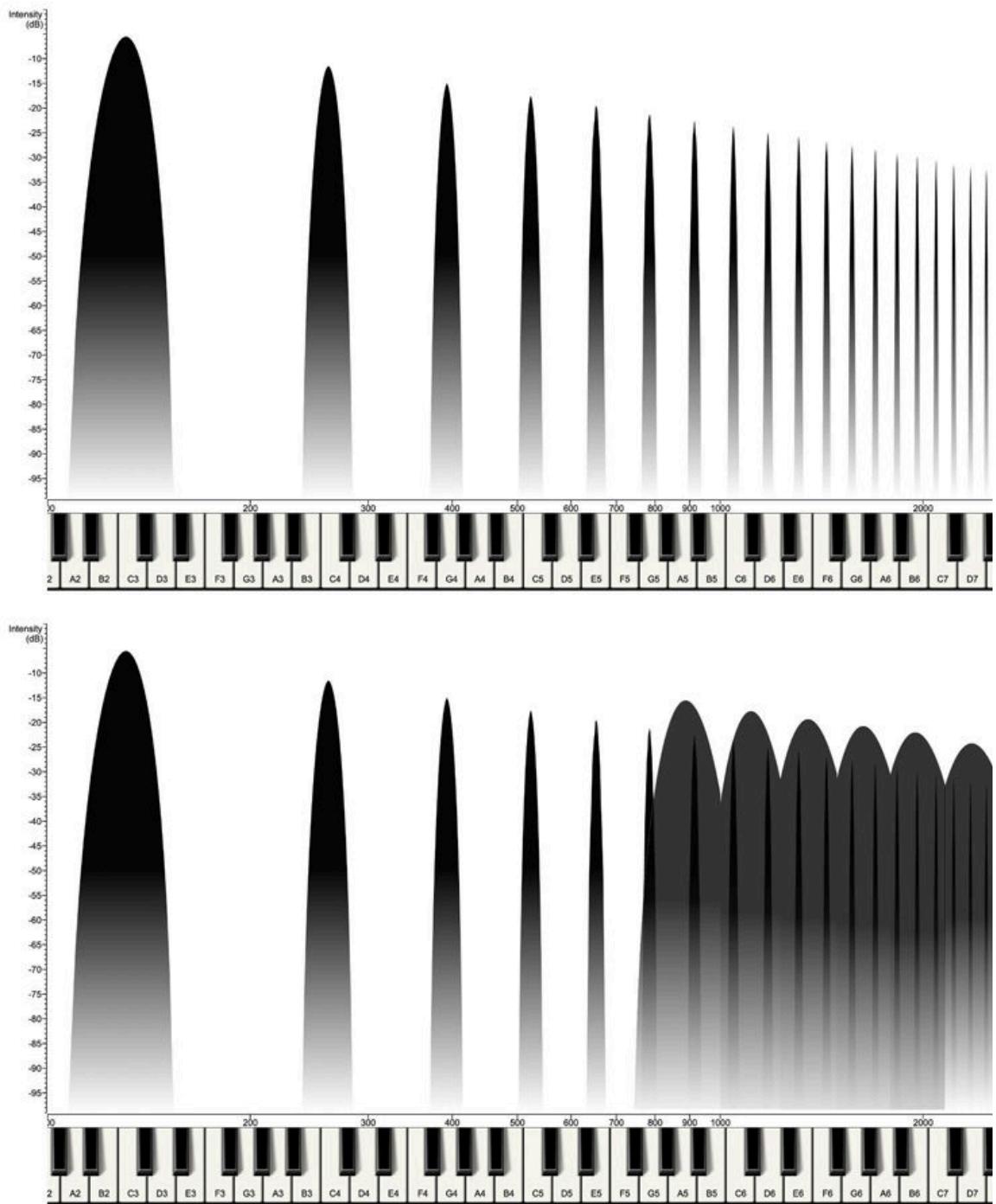


Figure 7.3 Power spectrum of a sawtooth wave at the pitch B2 (top). With schematic critical bands overlaid $> 5f_0$ (bottom). Source: Synthesized by author in Madde.

THE EQUIVALENT RECTANGULAR BANDWIDTH

A strict definition of auditory roughness suggests that at low frequencies, critical bands are wider in terms of musical intervals and narrower in terms of the difference in frequency.¹⁷ , ¹⁸ Think of this as an inversely proportional relationship that does not matter for much of the pitch range singers occupy. The higher the frequencies of the competing tones, the narrower the musical interval and wider the difference in frequency. To account for the nonlinear response curve of the basilar membrane, we must turn to the *equivalent rectangular bandwidth* formula to discern whether any two specific tones will trigger auditory roughness. With a little care, this formula may be entered into a spreadsheet app, or you may calculate it manually. Howard and Angus attest that this formula works so long as the center frequency of the critical band (f_c) is expressed in kHz and falls between 100 Hz and 10 kHz.

$$\text{ERB} = 24.7 * ((4.37 * f_c) + 1) \text{ Hz}$$

Here are two examples where the ERB does a better job of explaining the presence of auditory roughness than either the musical interval or raw frequency difference between two tones: Pure tones at G2 (~100 Hz) and C3 (~130 Hz) will exhibit a similar roughness as pure tones at C6 (~1,059 Hz) and D6 (~1,189 Hz). In this example, solving for G2 (~100 Hz, or 0.1 kHz for the formula) gives a bandwidth of 35.5 Hz. As C3 (~130 Hz) is just within that range, the ERB formula predicts that it will trigger roughness. Listening to this dyad confirms it. Similarly, the solution given C6 (~1,059 Hz, or 1.059 kHz for the formula) is a range of 139 Hz. D6 (~1,189 Hz) also falls just within this range, which can be confirmed by listening. Notably, the same musical fourth between the G2 and C3 would trigger no auditory roughness between a C6 and F6. It is interesting that lower-sung fundamentals could conceivably produce harmonic energy capable of triggering the roughness caused by harmonics at G2 and C3 (which would have to be $3f_o$ and $4f_o$ of the pitch C1). Keep in mind that sung pitches this low are uncommon but not unheard of. See Lab #11.

Considering significantly higher pitches momentarily, an interdependent relationship with the 5 kHz pitch perception cutoff seems to emerge. In other words, even if harmonics greater than $5f_0$ no longer fell within the ERB, they would trigger a different kind of noise anyway. Let us explore this with an example: Plug the pitch F5 (~699 Hz) into the ERB formula, and it returns a bandwidth of 402 Hz for the fifth harmonic ($5f_0$). Because the sixth harmonic ($6f_0$) must be 699 Hz faster than ($5f_0$), auditory roughness should be reduced in this part of the harmonic series at this and higher fundamentals. However, the energy represented by those higher harmonics soon crosses the upper threshold of pitch perception itself (around 5 kHz). *You absolutely must listen to this for yourself.* I certainly perceive a difference in quality between lower and higher harmonics in this pitch range, and the same buzzy label applied to lower pitches makes sense here as well, even if it occurs by a different mechanism. This is an underexplored area of perception, and I hope that more work will be done to identify these thresholds. See Lab #12.

In [chapters 3](#) and 6, I touched on the idea that the persistent energy we see in a spectrum or spectrogram at integer multiples of the fundamental is not the same thing as a chord of pure tones from independent sources. Those independent pure tones would each be a separate, persistently oscillating fundamental. Remember, harmonics are a way to understand the complexity of a repeating pattern. Harmonics may represent persistent oscillations of pressure that carry over from period to period or pressure changes that quickly dampen per glottal cycle. I encourage you to be open to the apparent paradox in this model. Even if the harmonic view clearly predicts auditory roughness, it may do so because it effectively summarizes the complex pattern that stimulates the basilar membrane. It is the response of the basilar membrane to that complex stimulus that generates this percept. We do not need to conceive of harmonics as functioning in a way that is equivalent to separately produced pure tones at the same frequencies, even though it is convenient to model them this way.

So, what voice functions *should* trigger auditory roughness? Or put another way, for what aspects of singing might auditory roughness act as

a *confirmation* that the function has been achieved? This could be a brighter vowel, a heavier registration, an increase in the vertical contact depth of the vocal folds, a narrowing in the vocal tract, an acoustic tuning that assists the supraglottal impulses to drop in pressure more quickly, moving from breathy to flow phonation or from flow to pressed phonation. Noisier and more extreme ways of singing will leverage this percept as well. As a reasonable rule of thumb for reading a spectrum, anything that brings more energy into the spectrum of a periodic sound above the fifth harmonic ($5f_0$) will trigger auditory roughness. In practical terms, in normal phonation *anything that increases the speed of the supraglottal pressure drop per glottal cycle and allows pressure changes much faster than $5f_0$ in the vocal tract will increase energy in the brighter part of the spectrum and likely trigger auditory roughness.*

Think of the buzziness of auditory roughness as the cognitive experience of a sound the ear cannot fully interpret. Auditory roughness is the sound of aural confusion. It is the sound of stimulating multiple frequencies within a *critical band of hearing*. In the context of how slippery a percept timbre can be,¹⁹ the presence of auditory roughness dependably signals which specific kinds of spectral energy are present.

THE INTERSECTION OF PITCH RESOLUTION AND BUZZINESS IN A SPECTRUM

Auditory roughness then relates to voice type, pitch, vowel, and intensity only insomuch as these variables affect the intensity of harmonics higher than $4f_0$. Our rule of thumb suggests that from the fifth harmonic and higher, all harmonics of a human voice fall within a critical band (in other words, a minor third) of a neighboring harmonic (see [figure 7.4](#)). Sundberg notes that the higher the amplitude of such harmonics, the stronger the auditory roughness they contribute to a singer's timbre.²⁰

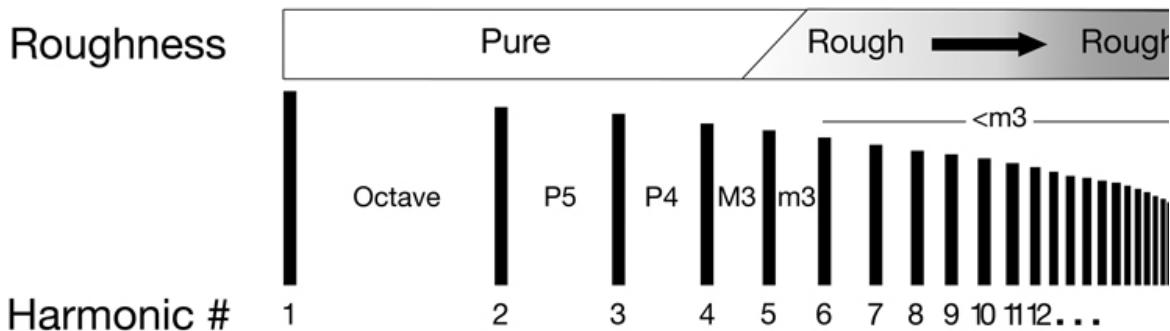


Figure 7.4 A generic harmonic series. Note that the musical intervals formed by the harmonics diminish in size as one moves higher in the series. From the fifth harmonic and higher the interval is small enough to elicit auditory roughness. Source: Author's schematic.

Pitch percept and auditory roughness are therefore interrelated. The unresolved (pitch-less) portion of the spectrum (generally from the ninth harmonic and above) will always elicit auditory roughness (see the portion of figure 7.5 labeled "Rough and Unresolved"). Harmonics five ($5f_0$) through eight ($8f_0$) will resolve into the pitch and *also* trigger auditory roughness (see "Rough and Resolved" in figure 7.5). Harmonics one ($1f_0$) through five ($5f_0$) will resolve into the pitch, but they will not trigger auditory roughness (see "Pure and Resolved" in [figure 7.5](#)).

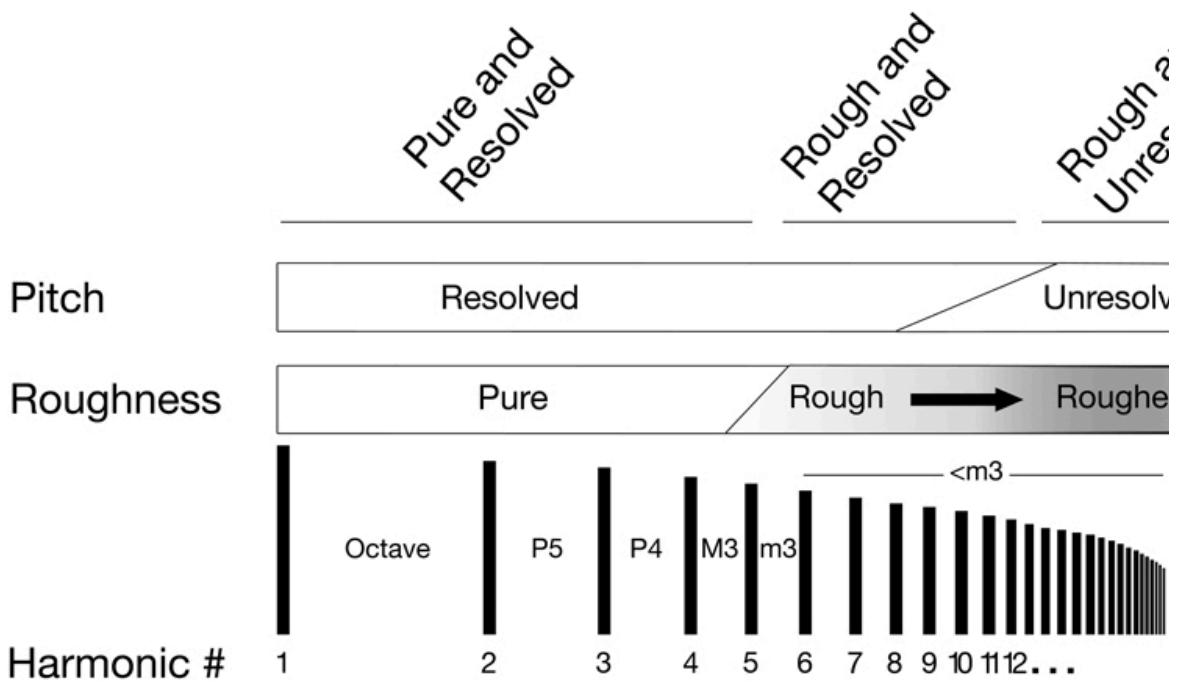


Figure 7.5 A generic harmonic series showing the broad groups of harmonics by the intersection of pitch resolution and auditory roughness. Harmonics 1–5 ($1f_0$ – $5f_0$) are pure and resolved. Harmonics 5–8 ($5f_0$ – $8f_0$) are rough and resolved. Harmonics 9 and higher ($\geq 9f_0$) are progressively rougher and unresolved. Source: Author's schematic.

THE TWO-TIMBRE MODEL AND BEYOND

By conceptually dividing the spectrum into a pure portion and a buzzy portion—and by further anticipating that the higher-frequency parts of that buzzy portion trigger pitch-less, noisy qualities—we can now map these ideas onto our familiar models of voice register and quality. This allows us to note qualitatively-opposable percepts in the otherwise prohibitively-complex experience of timbre. My aim is to answer Donald Miller's (2008) call to "free . . . hearing from the prison of subjectivity,"²¹ albeit without ultimately relying on the real-time visual feedback of a spectrum or spectrogram. To do so, let us first explore how these percepts associate with qualitatively opposable vocal functions.

It is challenging to try to discuss registration and registers in print. For the next image, I will use shorthand to indicate the difference between what most singers will understand to be their chest register (their speaking-pitch range) and their falsetto or head register (their weaker upper pitch range). Herbst and Švec (2014) suggest that the former is characterized by engagement of the *thyroarytenoid* (TA) muscles that make up the body of the vocal folds and the latter is characterized by the relative lack of that muscular engagement.²² The nature of the glottis is also affected by the engagement of the muscles that position the arytenoid cartilages at the posterior end of the vocal folds (the *-interarytenoid*, *lateral cricoarytenoid*, and *posterior cricoarytenoid* muscles), and the muscles that help to tilt the *thyroid* and *cricoid* cartilages to stretch the vocal folds (the *cricothyroid* muscles). But TA+ (*thyroarytenoid* engaged) and TA- (*thyroarytenoid* not engaged) point well to the gross role of the *thyroarytenoid* muscles in determining what might be called "laryngeal register." These terms likely do not map intuitively to every singer's sense of their singing registration, which is generally not binary in a practical sense. However, most people can yodel and notice the qualitative difference below and above that flip.

Figure 7.6 shows schematic spectra of the theoretical spectral slopes of the vocal fold impulse train in falsetto (TA-) and chest (TA+) registers

based on thyroarytenoid engagement. Note that relative to falsetto register, chest register has more energy in higher harmonics. The pure and buzzy regions are overlaid. Notice that buzziness is a more prominent perceptual feature in chest than in falsetto. Figure 7.7 similarly displays voice source qualities as discussed by Sundberg with the pure and buzzy regions overlaid.²³ This suggests that one practical way of identifying these voice qualities may be in terms of the ratio of pure and buzzy. Figure 7.6 suggests that chest is characterized by the presence of buzziness. Compare figure 7.7 to figure 7.6. Note that moving from breathy phonation to flow phonation similarly adds more buzzy, higher harmonic energy, while moving from pressed phonation to flow phonation adds pure warmth.

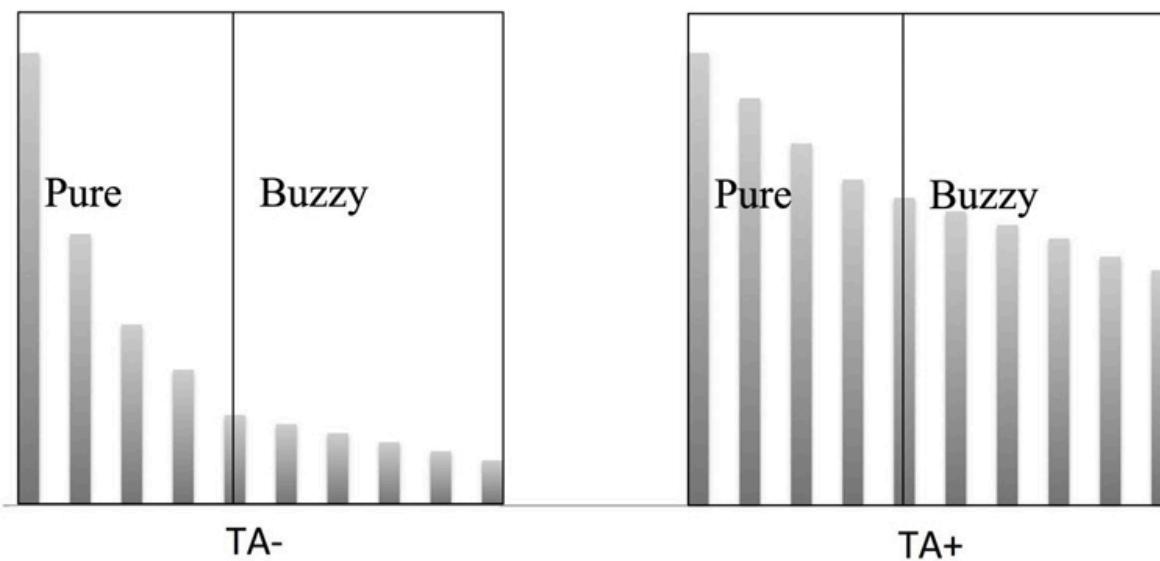


Figure 7.6 Schematic of head (TA-) and chest (TA+) registers by thyroarytenoid muscle contraction with pure and buzzy perceptual regions indicated. Source: Author's schematic.

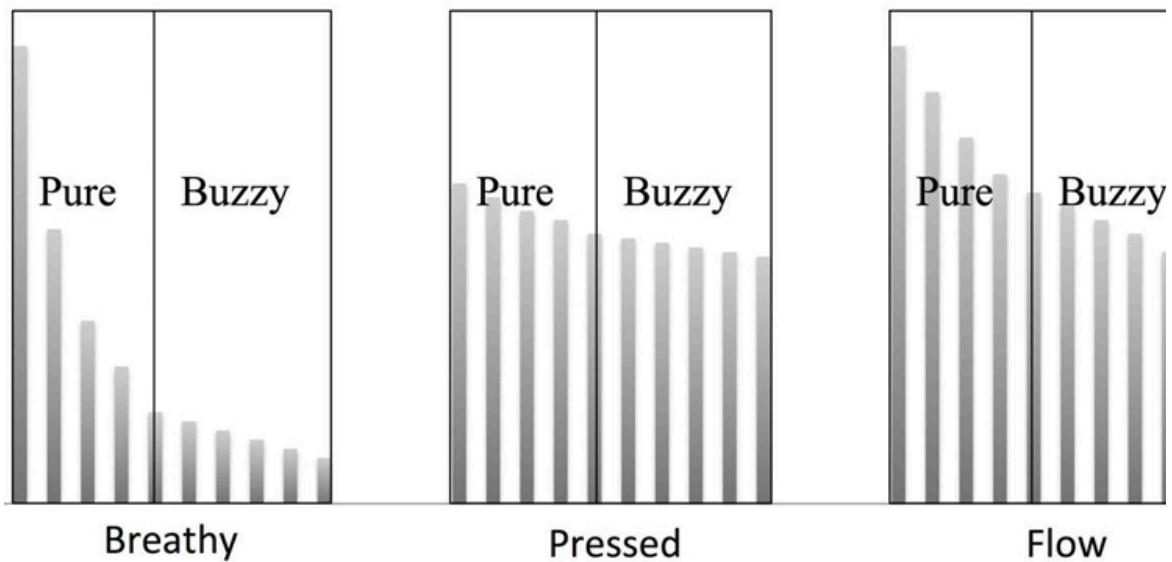


Figure 7.7 Schematic of spectra of voice source qualities with pure and buzzy perceptual regions indicated. Source: Author, based on Sundberg, *The Science of the Singing Voice*, 79.

The theoretical transglottal airflow patterns associated with these three phonation qualities are illustrated in figure 7.8. Breathy phonation features a high overall airflow rate that never reaches zero airflow. This is true even during the contacting portion of the breathy glottal cycle. The drop in flow shown at the beginning and end of the cycle is slow for breathy phonation relative to the flow and pressed qualities. This means the acceleration of the drop in pressure of the supraglottal air mass is slower for breathy phonation, which is why the faster (higher frequency) components present in pressed and flow are missing from that spectrum. Any brightness in breathy phonation tends to be stochastic turbulence. The flow and pressed patterns share the faster drop in airflow as the folds contact, but the total volume of transglottal airflow is much higher for flow phonation than pressed.

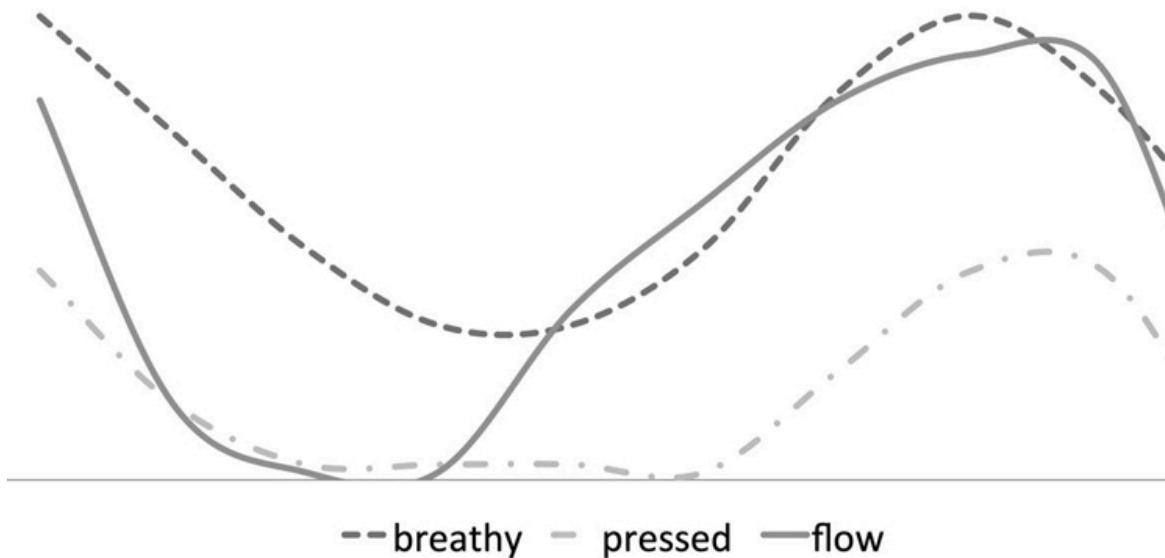


Figure 7.8 Transglottal airflow patterns of the three voice qualities from figure 7.7. One glottal cycle is shown here. Source: Author after Sundberg, *The Science of the Singing Voice*, 85.

Keep in mind that these patterns are the result of the interplay of tracheal pressure, glottal resistance, the various adjustments of the vocal folds, and the flow resistance of the vocal tract. Also keep in mind that this shows airflow patterns that facilitate the supraglottal impulse—the interaction of both laryngeal and vocal tract dynamics—rather than that impulse itself. Still, key features of [figure 7.8](#) should map well to the source impulse spectra in figure 7.7.

These three voice qualities are interrelated functionally: Breathy and flow feature high transglottal flow volumes; flow and pressed feature a rapid cessation (or decline) of flow. This suggests that they may be used to bridge the gap between workable and unworkable functions, especially if we replace the somewhat pejorative terms *breathy* and *pressed* with a continuum of vocal fold adduction. This adduction may be caused by muscular effort, aerodynamic forces, or acoustic forces.

I think it is then an attractive notion to float a perceptual definition of chest voice and falsetto (or head voice) based on the presence of auditory roughness. Likewise, *mixed* registers may be at least partially explained in terms of these perceptual qualities (pure/buzzy) rather than the physiological consistency of their production. This suggests that “mix” is

the presence of brightness at effort levels below that found in full, speech-pitch-range singing. “Seamless” register transitions may be explained in terms of incremental changes in the presence of auditory roughness as pitch rises or falls. We can tilt toward the stereotypical perceptual signifiers of either falsetto or chest voice anywhere within our singing range by changing our phonation quality or vocal tract shape to rebalance the ratio of pure and buzzy. For example, the use of bright vowels like /e/ or /æ/ in belt and mix-belt registration facilitates the impression of chest voice’s buzzy brightness without the TA engagement that would only work in a lower pitch range.

Changes in dynamics also impact timbre. [Figure 7.9](#) shows the waveform and three spectra of a cisgender man singing a typical *decrescendo*. Each spectrum falls under the time range it represents. Notice that the higher frequency parts of the spectrum decrease in intensity faster than the lower frequency parts. The levels of $1f_0$ and $3f_0$ are indicated in each spectrum as a reference. Notice as well that the brighter parts of the spectrum—the portion that will elicit auditory roughness—drops in intensity faster than the warm part of the spectrum does. Titze (2000) suggests that many things might impact vocal intensity: adductory forces of the vocal folds, air pressure below the glottis in excess of the phonation threshold pressure, formant tuning (acoustical relationships between the vocal folds and vocal tract), general vocal tract adjustments to favor higher frequency energy, or fundamental frequency.²⁴ Each situation sounds a little different, but except in cases of very high fundamental frequencies, a louder version of the same sound will tend to feature more auditory roughness.

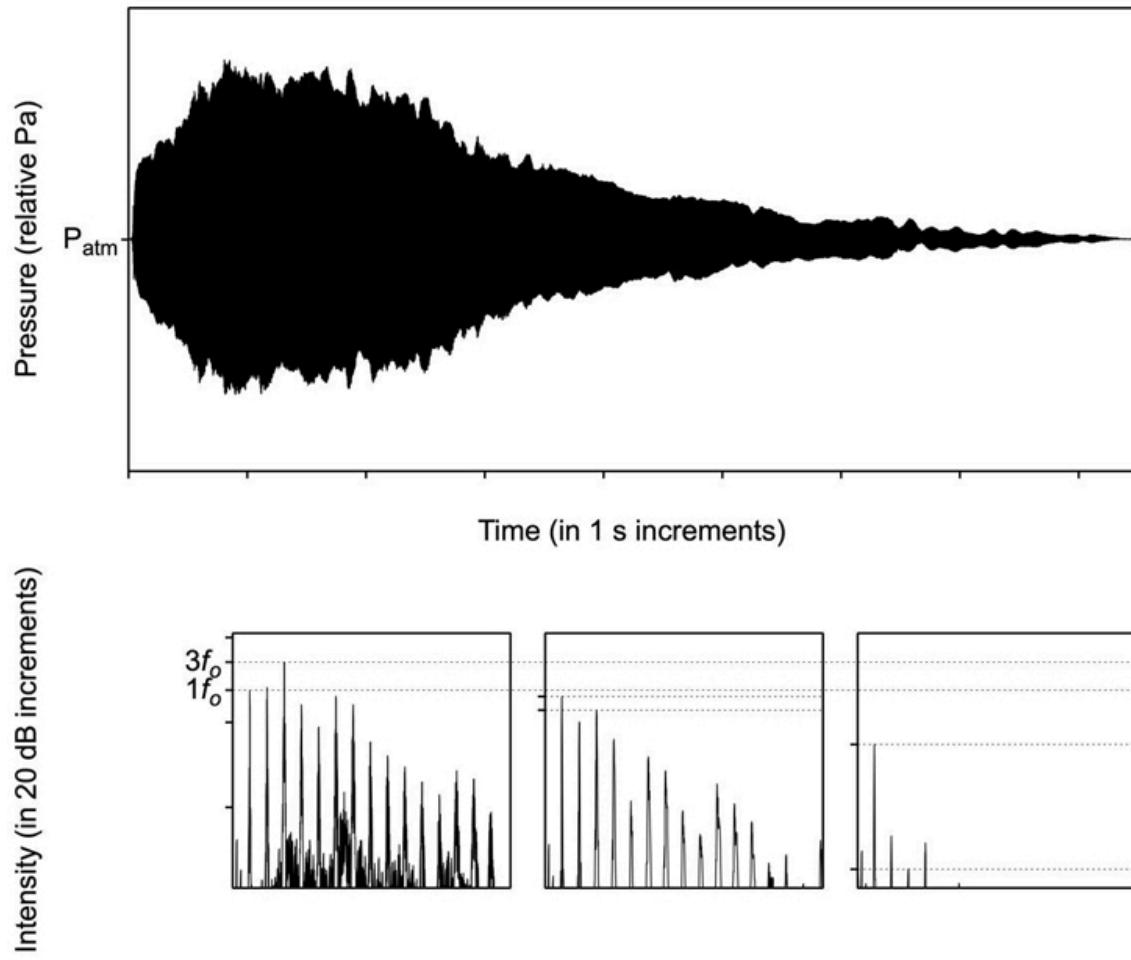


Figure 7.9 A cisgender man executing a typical decrescendo. Waveform (above) and three spectra (below) corresponding to the time range each aligns with. Source: Recorded by the author under controlled conditions.

POTENTIAL CONCLUSIONS AND APPLICATIONS

In order to understand the vocal function of a singer, we must be able to hear the effects of the various physiological and aerodynamic coordinations that generate the sound. Auditory roughness might appear to be a broad idea, but in periodic voicing, auditory roughness arises when sufficiently fast pressure changes occur repeatedly. Auditory roughness is a cue that *something* is causing that fast, repeating pressure change. It could be a sign of flow phonation paired with a supraglottal vocal tract air mass ready to quickly snap from high to low pressure. It could be a sign of increased vocal fold mass thickness—even to the point of losing balanced phonation. It could be a sign that the resonances in the vocal tract are timing particularly well with each vocal fold contacting event. Auditory roughness could point to deeper contacting of the vocal folds per glottal cycle (typically associated with more TA engagement),²⁵ or it could point to the complete closure of the arytenoid cartilages in the absence of significant TA engagement. Or it could be one of several adjustments that incrementally increase or decrease the intensity of the voice. As I wrote above, auditory roughness tells you about the speed of pressure changes in the vocal tract; as such, it is a broadly useful percept when combined with other perceptual qualities and contextual clues.

In spectra and spectrograms, we see various spectral structures in different groups of higher or lower intensity harmonics. Different portions of a vowel's spectrum have historically been assigned different roles. The lowest two to three peaks in the spectrum are thought to establish vowel identity. Higher peaks define brightness. The intersection of auditory roughness and pitch resolution cuts across this division and suggests that any peak can also be either pure, buzzy, or noisy, depending on pitch and the frequencies of the harmonic components of the peak itself. At least theoretically, it is possible that a sung pitch could be so low that all vowel-defining peaks would be buzzy, and that the majority of the spectrum would remain unresolved from the pitch. If you consider the sound of a very low bass's speaking voice, buzzy noise is the defining characteristic.

It is also possible to sing a pitch so high that any very little audible buzzy sound exists. So, apparently similar spectral features do not automatically indicate similar percepts.

Perhaps the most important implications of auditory roughness and pitch resolution relate to the subtle rebalancing of multiple functional variables that characterize registration. The voice pedagogy community uses a variety of terms to explore both raw laryngeal mechanisms and the endless number of *in-between* settings. The buzzy quality of auditory roughness or pitch-less noise may meaningfully differentiate currently observed but difficult to nail down vocal qualities. Consider that changes in dynamics, vowels, vocal tract shape, air pressure, transglottal airflow, glottal adduction, effort, emotion, and phonation quality all impact the spectral slopes (the relative strength of slow and fast pressure changes) of the glottal impulses and the radiated sound. Rather than imagining these to be descriptive, analytical terms, think of them as outcomes to help shape the simple prompts you offer. They are *levers* that can effect change in a voice, in a way of thinking. If these adjustments modulate the amount of energy above the fifth harmonic ($5f_o$) shown on a spectrum or spectrogram, auditory roughness is the perceptual correlate to that physical phenomenon. Strong energy even higher in the spectrum becomes pitch-less noise.

DISCUSSION QUESTIONS

- The term *roughness* means different things to different communities. Name a few of these different uses and consider what they may have in common.
- With isolated pure tones we may explore the phenomenon of *beating*. This occurs when those two tones are closer in frequency than 15–20 Hz. What pitch would a human have to sing for higher harmonics to elicit this kind of beating?
- Expressed in a musical interval, what is the rule of thumb to determine if two adjected harmonics will trigger auditory roughness.
- That rule of thumb only describes a portion of the sung range. Would we expect that musical interval to get bigger or smaller as pitch lowers below C2?
- Why would pure tones at 100 Hz and 106 Hz beat at the same rate as tones at 200 Hz and 206 Hz, while those dyads represent different musical intervals?
- Explain *difference tones*. Would you expect the difference tone to raise or lower in pitch the closer two pure tones are in frequency?
- What is the utility of the *equivalent rectangular bandwidth* over the musical interval rule of thumb for predicting auditory roughness? What cases does it do a better job of explaining?
- Describe a few contrasting voice qualities or registers in terms of the auditory roughness that should be present or absent.
- In physical terms, what kind of pressure change must exist somewhere in the voice signal to generate auditory roughness? Can you name a few ways this kind of pressure change might be generated in a voice?

NOTES

1. Eric J. Heller, *Why You Hear What You Hear* (Princeton, NJ: Princeton University Press, 2013), 508.
2. Johan Sundberg, *The Science of the Singing Voice* (DeKalb: Northern Illinois University Press, 1987), 108.
3. Ingo R. Titze, "Why Do Close Harmonies and Dissonances Sound Rougher at Low Pitches than High Pitches?" in *Journal of Singing* 73, no. 4 (March/April 2017): 411–12.
4. Ian Howell, "Necessary Roughness in the Voice Pedagogy Classroom: The Special Psychoacoustics of the Singing Voice," *VOICEPrints* (May/June 2017): 4–7. Portions reprinted with permission.
5. "Voice Qualities," The National Center for Voice and Speech, <https://ncvs.org/voice-qualities/>.
6. Mathias Aaen, Cathrine Sadolin, Anna White, Reza Nouraei, and Julian McGlashan, "Extreme Vocals—A Retrospective Longitudinal Study of Vocal Health in 20 Professional Singers Performing and Teaching Rough Vocal Effects," in *Journal of Voice* (article in press, 3 June 2022), DOI: 10.1016/j.jvoice.2022.05.002.
7. This framing is attributed to Kenneth Bozeman in lectures.
8. David M. Howard and Jamie Angus, *Acoustics and Psychoacoustics*, 5th ed. (New York: Routledge, 2017), 84–85.
9. Eric J. Heller, *Why You Hear What You Hear* (Princeton, NJ: Princeton University Press, 2013), 508.
10. Heller, *Why You Hear*, 509.
11. Sundberg, *The Science*, 109.
12. Leila Chaieb, Caroline Wilpert, Thomas P. Reber, and Juergen Fell, "Auditory Beat Stimulation and Its Effects on Cognition and Mood States," in *Frontiers in Psychiatry* 6 (2015): 70, <https://doi.org/10.3389/fpsyg.2015.00070>.
13. Howard and Angus, *Acoustics and Psychoacoustics*, 85.
14. Sundberg, *The Science*, 109.
15. The rule of the minor third really only applies to frequencies higher than ~523 Hz (C5).
16. David M. Howard, "Hearing Modeling Spectrography," in eds. Graham F. Welch, David M. Howard, and John Nix, *The Oxford Handbook of Singing* (Oxford: Oxford University Press, 2019), 1077–79.
17. Pantelis N. Vassilakis, "Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance" (PhD diss., University of California, Los Angeles, 2001), 194.
18. Howard and Angus, *Acoustics*, 85.

19. Cornelia Fales, "The Paradox of Timbre," in *Ethnomusicology* 46, no. 1 (2002): 58.
20. Sundberg, *The Science*, 108.
21. Donald G. Miller, *Resonance in Singing: Voice Building through Acoustic Feedback* (Princeton, NJ: Inside View Press, 2008), 46.
22. Christian Herbst and Jan Švec, "Adjustments of Glottal Configurations in Singing," *Journal of Singing* 70, no. 3 (January/February 2014): 302–3.
23. Sundberg, *The Science*, 85.
24. Titze, *Principles*, 251–59.
25. Titze, *Principles*, 291.

8

Tone Color and Brightness

This chapter addresses the perceptual continuum of *tone color*, a term I will narrowly define in the hope of drawing attention to a specific aspect of timbre. Tone color offers us a way to understand qualitative oppositions along the scale of frequency. Tone color points to the idea that pure tones at different frequencies are qualitatively different, and that pure tones at the same frequency are qualitatively similar. These pure tone qualities in turn *inform* the frequency content-dependent tone colors that are present in complex tones, such as a sung vowel. This is to say that complex tones are characterized by both faster (high-frequency) and slower (low-frequency) pressure oscillations that occur simultaneously, and that these different pressure change aspects of a complex tone trigger different tone color percepts. In real-world terms, this pressure change information is generated in a singing voice by both the number of repetitions of the complex resonant response of the vocal tract per second (fundamental frequency) and the recurring existence of faster oscillations within that complex pattern (the resonances themselves). Tone color characterizes our perceptual response to the *speeds* of all of these pressure changes.

In [Chapter 6](#) I suggest the utility of thinking of resonances as a time-and-pressure domain phenomenon based on the excitation patterns of the vocal tract air mass, rather than as the latent amplification potential of the vocal tract. The reader will recall that the source-filter model suggests separating the contribution of the vocal folds and vocal tract, and further makes distinctions regarding the spectral details of the radiated sound.

This leads to a set of model-specific terms to label each part, wherein the vocal folds produce harmonic frequencies ($1f_o$, $2f_o$, $3f_o$, etc.) which are amplified (or not) by the vocal tract resonances (f_{R1} , f_{R2} , f_{R3} , etc.). The combination of those vocal fold harmonics and vocal tract resonances produces peaks in the spectrum which may be labeled *formants* (indicated with F_1 , F_2 , F_3 , etc.).¹ Thus, if a harmonic has no available resonance, that harmonic will be attenuated in the radiated spectrum. If a resonance has no harmonic to populate it, it cannot turn into a formant and functionally will not exist.

The history of the term *formant* is worth a brief mention. In 1894 Ludimar Hermann was studying phonograph recordings of sustained vowels.² When he visually amplified the grooves in those phonograph discs, he saw the same familiar periodic excitation and damping waveform patterns we can now see with modern computers. Hermann applied a Fourier transform to these repeating patterns, representing the repeating oscillations of the vocal tract resonances as spectral peaks. He termed this spectral representation of the time-and-pressure domain phenomena *formants*.³

I offer this brief history now because the material covered in this chapter will reference both the speech literature and the seminal work of singing voice researcher Donald G. Miller. This combined body of literature uses the term *formant* to describe those radiated peaks and frequently frames observable phenomena using terms and perspectives from the source-filter model. For example, consider this common construction: Align $1f_o$ with f_{R1} to produce a strong F_1 . This chapter will use the term *formant* to align with these sources. But please keep in mind that we know that the feature of a spectrum labeled as a formant is generated mathematically by averaging the periodic repetition of the excitation and damping patterns of the vocal tract per glottal cycle over time. What we are calling a formant is a spectral feature that points to the physical oscillations of pressure within the vocal tract. Consider that $1f_o : f_{R1}$ produces F_1 can also be described in terms of timing each subsequent

glottal contacting event with the end of the first complete oscillation of the slowest resonance in the vocal tract.

Since we understand that these terms all exist within models, and that some models need labels for specific concepts, I think it is possible to use the term *formant* in this chapter without pointing to an abstract concept. The spectrum typically averages away short-time fluctuations in the amplitude of the vocal tract resonances, representing them instead as continuously intense harmonics. The human ear is generally stimulated by that complex pattern, not continuous harmonic frequency energy as it appears on a spectrum. However, summarizing these damped oscillations in this manner allows one to talk about their frequency in practical terms and with great specificity.

Depending on what you have read, you may conceive of a formant as the result of combining a harmonic with a resonance, or you may conceive of it as a spectral average of multiple periods of the damped excitation pattern of the vocal tract. Or you may have never grasped the idea of a formant. A gentle reminder that it is probably best to not get locked into just one way of understanding the phenomenon these terms point to, or to imagine that the abstract concept captured by the term *is* the underlying phenomenon.

WHAT IS BRIGHTNESS?

In the voice pedagogy and vocology literature, any discussion of frequency tends to reference values in cycles per second, or hertz (Hz). For example, a soprano singing an A5 may generate a strong fundamental at 880 Hz. A tenor's voice may feature a prominent cluster of resonances around 2,900 Hz. A rock singer may have persistent energy between 6 kHz and 12 kHz. But how do these frequency ranges *differ*? Using objective measurements is a simple and granular way to discuss frequency, but it is one devoid of qualitative information. For example, is 1,000 Hz as different in brightness from 500 Hz as 2,000 Hz is from 1,500 Hz? Can meaningful comparisons be framed in these terms? Those who grapple with these questions tend to use the term *brightness* to discuss the frequency spectrum. Frequency tells us something measurable about a repeating pattern. Brightness honors the perception that high-frequency tones are characteristically different from low-frequency tones.⁴

A deeper understanding of brightness would acknowledge qualitative oppositions between low- and high-frequency pure tones. It would also suggest that two pure tones that share a common frequency are similarly bright. This does not suggest that sounds with the same pitch are automatically similarly bright. Nor does it suggest that two sounds with a similar spectral peak will sound the same overall. Likewise, two pure tones need not share all aspects of timbre to share brightness. For example, perhaps only one source has vibrato or amplitude pulsing. However, a stimulus at a given frequency will generally elicit a consistent brightness percept, regardless of source.

Perhaps the strictest view of brightness would suggest that anything more complex than a pure tone can be understood through this principle, but only with caveats. I suggest that the various spectral peaks within a single, steady-state vowel can be considered to have different *brightnesses*, but this assumes one has chosen to attend to the sound in detail in the first place. Even the lowest two spectral peaks—the vowel formants (F_1 and F_2)—can be thought of as each having a certain

brightness, although the short syllable duration and relatively low pitch of speech obscures such details. The brightness of a vowel formant is typically less bright than that of higher frequency energy, but it is brighter than lower frequency energy. How perceptually different such peaks may be from one another depends on how near or far they are from one another on the brightness scale. In the percept of the listener, those peaks may interact in complex ways, but they coexist along the same perceptual continuum as more obviously bright and dark pure tones. This is all to say, *the colors we perceive to be present in a voice are all shades of tone color on the same continuous scale of brightness that characterizes any sound.*

What follows explores the long history of these ideas, establishes limits on the extent of this phenomenon, explicates a framework for encountering the tone colors of pure tones, and explains how one might extrapolate the multiple tone colors present in complex tones.

TONE COLOR: CONFLICTS BETWEEN SPEECH AND MUSICAL SOUNDS

Tone color is perhaps the most controversial idea covered in this book. It is easy to experience but challenging to write about. I would like to suggest that part of the potential resistance to tone color as a phenomenon is semantic; another part is due to a lack of grounding in historical context; and much of the remaining resistance represents basic conflicts between those who study speech and language and those who study the timbre of musical sounds (including singing). No one who studies the voice argues against the idea of co-presenting sound colors. This idea is already encoded into our use of the term *chiaroscuro*, or “bright darkness.” A complex sound can have multiple aspects of brightness and darkness at once.

One frequently finds statements in the voice pedagogy literature to the effect that vowels have inherent pitches,⁵ or that what differentiates one vowel from another is the presence of higher or lower frequency spectral peaks. William Vennard (1967) notes that certain vowels also *share* spectral peaks:

When one sings Ay [e], he is really singing Oh [o] plus a high partial which is not heard in the Oh [o]; and when one sings Ee [i], he is really singing Oo [u], plus a still more ringing overtone.⁶

Roughly put (and this is not strictly accurate), adding higher frequency energy to [o] and [u] results in [e] and [i]. I urge you to hear this and experience it personally in Lab #13. It is challenging to accept this without wondering if what characterizes those brighter vowels exists primarily in that higher frequency energy. This is a departure from the prevailing idea in speech research that vowel identification and paralinguistic cues about the speaker leverage the relationship between the frequencies of the formants rather than the frequencies of the formants themselves.⁷ This view is called *formant-scaling*.

The common formant-scaling argument is that the same phoneme uttered by different speakers—or by the same speaker attempting to imply they are physically larger or smaller than they are—may be understood based on the rough similarities of their formant structures. This is despite sometimes radically differing vocal tract dimensions, accents, languages of origin, and objective formant frequency measures. This is a reasonable way to simplify the complex variability of speech sounds between speakers. Evidently, the perceived vowel is the entire formant structure all at once. How else could speakers with differing formant frequencies convey the same vowel?

However, singers are arguably as interested in the aesthetic qualities present in a sound as they are in the linguistic content of that sound. Grafting this framework on the study of the singing voice—where the vowel is “all of those formants at once”—may lead to analysis paralysis. However, if this were true, how could treble singers make vowel sounds at pitches higher than the first formant?

The answer to this dilemma seems reasonably simple. In this context, the idea of a *vowel* does not point to a specific *sound* so much as it points to a range of sounds based on the behavior of the articulators in the context of the surrounding phonemes found in spoken language. The notion of a vowel—and of a *phoneme* more broadly—points to a group of somewhat similar sounds that all serve the same function in spoken language. This is because the idea of a *vowel* within such a model is conceived semantically; the vowel is less tied to the timbre of the sound than it is linked to the meaning that sound evokes in a linguistic context. A vowel is really a conceptual “container” for a group of differently timbred sounds. So we would expect that the timbres of a child and an adult speaking the same vowel would be dissimilar, even if their articulators trace similar paths relative to their anatomy. Likewise, the timbre of someone speaking an American [a] with a thick Russian accent is different from that of a native speaker of American English from the Midwest.

Additionally, relative to singing, speech generally occupies a restricted and lower average pitch range. Titze (2017) writes that “the larynx as a

sound source in speech is limited in its range and capability because a low fundamental frequency is ideal for phonemic intelligibility and source-filter independence.⁸ If speech occupies a lower pitch range than singing, this means the vocal tract is allowed to resonate longer for each pitch period. This longer period more fully realizes the complexity of the resonant response of the vocal tract, which in turn allows more subtle variability in timbre. As Titze writes, this also increases source-filter (vocal folds and vocal tract) independence as the amplitude of the oscillations in the vocal tract has time to significantly dampen between glottal contacting events. These time-and-pressure properties of speech appear in a spectrum or spectrogram as multiple harmonics that occupy the frequency ranges of each vocal tract resonance. As a result, a wide range of resonant formant frequencies are possible.

Within the context of the flow of spoken language—where ongoing timbral variation is a feature—the actual tone colors present in a spoken vowel do not tend to matter more than that vowel’s adjacent consonants, vowels, glides, or pauses. This process is called *coarticulation*.⁹ Denes and Pinson (2015) write that “the speech wave has very few segments whose principal features remain even approximately static.”¹⁰ For example, native English speakers will notice the difference between the /k/ sounds in *keep* and *cool* to get a sense of how the vocal tract shape of the vowels [i] and [u] change the tone color of the initial consonant. Now compare the [i] in *tea* and *peep* to note how the [i] is subtly changed by the initial consonant.

Singing is different. What is a *vocalise* or *melisma* but an exploration of the language-free timbre of a specific registration setting through a specific vocal tract shape at a specific pitch? Entire pedagogic approaches are built on the idea that vowels modify or migrate.¹¹ Singers cover to the top, narrow in the *passaggio*, bloom, add hoot, track ring or cut, align all their vowels, sing with a sob or cry, and so on. None of these descriptive terms are relevant to discussing functional speech or linguistic meaning. In many cases we may do better to consider the sound of a voice as we would a musical instrument. There are ways to *play* the voice that allow for the efficient and inefficient navigation of certain pitch ranges. We

sometimes color the voice in the process of achieving a technical outcome.

I started this chapter suggesting that tone color is a challenging topic to discuss with the necessary nuance. In this section I have pointed to challenges inherent in grafting conclusions from the study of speech onto singing. To continue our discussion of tone color, we return to the idea of *brightness*.

THE HISTORY OF BRIGHTNESS

Brightness as a parameter of sound is an old idea. As early as 1885, Ernst Mach suggested labeling low and high pure tones as *Dumpf* ("dull") and *Hell* ("bright" or "light").¹² Central to the idea of brightness is the notion that low- and high-frequency pure tones are *qualitatively* different. This idea may be extended to address how the faster and slower oscillations that compose complex tones may be thought of as dark or bright. This should not be controversial to anyone who has worked with singers. Some ways of singing sound warm. Others sound bright. Sometimes these different qualities are present simultaneously. Sometimes we must add brightness to a dark-sounding vowel to make it audible in certain performance contexts. What makes a pure speechlike [u] audible in a concert hall if not for that vowel being fortified with additional brightness that is unnecessary for speech? It makes as much sense to reduce brightness in a sound to the numerical scale of frequency in hertz as it would to do the same for visible light in lumens. While mathematically accurate, this is not how humans experience these phenomena.

Anecdotally, I have noticed movement in the fields of voice pedagogy and vocology toward describing the contributions of the first and second formants of the voice in looser, qualitative terms. Bozeman (2013) suggests that the first formant generally provides a "depth" and the second formant a "clarity" to a given vowel.¹³ This framing in perceptual terms is incremental but significant as we move toward assigning the perceptual qualities of brightness to the features we see on a spectrum or spectrogram.

Bozeman (2021) currently assigns the labels *under-vowel* and *over-vowel* to the tone colors of the first and second formants, respectively.¹⁴ These labels have found purchase within the voice pedagogy community, crystallizing the idea that one can hear simultaneously sounding qualities of a vowel without centering its semantic properties. As with all models, this one has its limits. Notably, the concept of an under- or over-vowel does not directly point to the changes in overall timbral complexity or tone

color that accompany rising pitch. Nor do these labels invite the potential for meaningful tone color contributions below and above the frequency ranges of the vowel formants.

In 2016 I similarly suggested that the first formant can broadly be labeled *warm*, the second formant *clear*, and higher formants increasingly *bright*, like the defining quality of an [i] vowel. Depending on the pitch and vowel, there may also be a further warm [u]-like quality from energy slower (lower in frequency) than the first formant. This slightly more nuanced model also allows one to notice timbral shifts related to pitch. For example, at the pitch F4 an [a] has an additional, darker [u]-like quality due to the frequency of the fundamental. An octave higher, and this quality disappears.¹⁵

Plomp (1966) shares a history of brightness that reaches back to the nineteenth century. It is worth including an extended passage here. I encourage you to seek out the original source, as it helps lay out the historical context for much of what follows:

Apart from . . . Mach (1885),¹⁶ Engel (1886)¹⁷ and Stumpf (1890)¹⁸ may be regarded as the first promoters of the idea that, in addition to pitch, simple tones have timbre (Stumpf's "Tonfarbe"). They realized that only on the basis of this assumption the timbre of complex tones can be understood.

Köhler (1909, 1911, 1915)¹⁹ came to the same conclusion. He gave attention to the fact, also observed by other investigators (Grassmann, 1877;²⁰ von Wesendonk, 1909),²¹ that simple tones have some resemblance, depending upon frequency, with particular speech vowels. Köhler (1911) established that, from low to high frequencies, the character of a simple tone shifts from the German /u/ over /o/, /a/, and /e/ to /i/. . . . In general the order of vowels attached to a tone moving from low to high frequency was confirmed (Weiss, 1920;²² Engelhardt and Gehrcke, 1930²³). It is plain that this resemblance is related to the location of the formants in the frequency range.

Recent research indicates that a better agreement can be achieved with two simultaneous simple tones (Morton and

Carpenter, 1962).²⁴ Since it is unlikely that the resemblance between simple tones and particular speech vowels is caused by pitch, it is, as Köhler rightly concluded, much more probable that simple tones are characterized by a distinct timbre.

As a name for this attribute of simple tones, the term *brightness*, already used by Mach as we saw above, appears to be the most appropriate one. . . . Although it may seem rather peculiar that two psychological parameters, pitch and brightness, are related to the single frequency parameter of a simple tone, this is not so strange if we accept that pitch is correlated with periodicity and brightness with frequency. For simple tones, period is the reciprocal of frequency, but this is not the case for complex tones.²⁵

Noteworthy in this survey laid out by Plomp is that generation after generation of researchers all observed the same basic principle: Pure sounds change tone color as frequency changes, and the way that such tone colors change bears a similarity to the way that vowels contrast as their formants change. Even Vennard suggests, one “whistle from bottom to top of your range and notice the change from ‘Who’ to ‘Whee’.”²⁶ This basic idea is more than a century old and has been reiterated by a respected research scientist as recently as Plomp (2002): “For the extreme case of a sinusoidal tone without harmonics, the tone’s frequency as its single variable determines pitch as well as timbre.”²⁷ Let us extend this logic to conclude that if all pure tones shared a timbre, Plomp would have no need to clarify this point. Another way to think about this is that vowels leverage this deeper perceptual response. **Tone colors are not vowel-like. Vowels are tone color-like.**

This idea appears in music-based theory and perception texts as well. Fritz Winckel (1967),²⁸ Wayne Lawson (1985),²⁹ and Robert Cogan (1998)³⁰ somewhat independently arrive at comparable conclusions. I arrived at the ideas that follow prior to my own exposure to Plomp, Winckel, and Lawson. Why then is this line of inquiry missing from modern voice pedagogy texts? Why is it missing from modern speech science and acoustics texts? It is certainly possible that we are all wrong,

suffering from a century-old hallucination at the intersection of neuroplasticity and suggestibility going all the way back to Mach. After you experience the sounds laid out in the labs, you may find yourself similarly hallucinating.

The framework I will describe in what follows is both flexible and limited. In the next chapter I do offer specific brightness labels that have survived extensive empirical exploration, but I am not dogmatic about them. I readily admit those labels have never been tested in the modern era with a large number of naive participants. Training subjects to limit their feedback to the relevant aspect of the stimulus would be onerous, and these percepts are liminal and hard to describe with words. Because some education is required before participating, this implies that this level of specificity is, to an extent, a learned behavior, much like the labels we use for the colors of visible light.

That said, the fact that this idea has repeatedly sprung up in various literature for almost a century and a half suggests to me that something important is at work. Well-designed, controlled experiments using modern methods should be carried out. They would undoubtedly spark even more interesting questions that are worth exploring.

WHAT IF FORMANTS HAVE TONE COLORS?

In 2013 I was a doctoral student working with the composer and music theorist Robert Cogan at the New England Conservatory. He was an early adopter of spectrograms in the analysis of structural design within music. His own work, influenced by generations of composers, focused on timbre as a compositional parameter in music³¹ and extended older ideas about tone color³² into the (then) modern age of computer-based spectral analysis.

Cogan thought in terms of two structural levels of tone color: First, that all frequencies (from low to high) may be grouped into grave, neutral, and acute colors.³³ Second, within those divisions lie subqualities evocative of several vowels. [Figure 8.1](#) illustrates this substructure, moving from the grave [u] and [o] into the neutral [a] and finally acute [e] and [i].

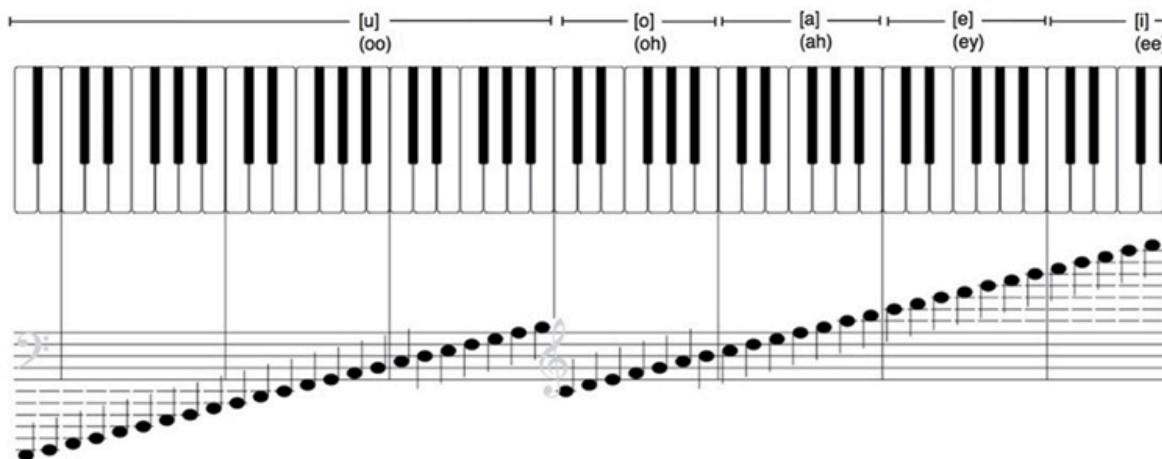


Figure 8.1 Robert Cogan's register-based analysis of the "sonic qualities" of sine waves. Source: A new schematic by the author based on Robert Cogan, *New Images of Musical Sounds*, pages 7, 12.

Note that Cogan uses these labels to group large swaths of frequencies, and that changes take place consistent with octaves starting on the pitch class C. This is a rough shorthand intended to allow a moderately granular analysis of timbre. While I do not think the fine details of this scheme hold up to scrutiny—for example, the strongest

vowel formant for [u] falls within the frequency range that Cogan suggests evokes an [o]-like color—I do want to assert that the concept being conveyed is important.

As I was exposed to these ideas, I was simultaneously immersed in a study of Donald G. Miller's book *Resonance in Singing* (2008). This seminal text in the application of acoustic theory and computer-based visual feedback to the act of singing is remarkably thorough in contextualizing the parameters of voicing investigated. It is notably devoid of detailed perceptual descriptions of sound beyond a single occurrence of the word *dullness* to describe a lack of second formant tuning in a tenor voice;³⁴ and the words *dark* and *darkness* to describe $1f_o : f_{R1}$ tuning of an endogenous estrogen puberty classical singer on the pitch F4.³⁵

Viewing Miller's book through the lens of Cogan's work suggests that while Miller repeatedly discusses the effect of a given formant tuning maneuver or effective closed quotient (now referred to as "contact quotient") on a singing voice, he does not discuss changes in the resulting sound with any granularity. He characterizes the quantitative acoustic output of singing but does not discuss how the ear is likely to respond to that output. In fact, he frequently conveys certain ideas by contrasting two differing sound samples. I think these explorations would be complemented by the simple, and fundamentally uncontroversial, *two-timbre model* introduced in [chapter 7](#).

Consider two acoustical relationships that Miller points to: *Hoot* and *Yell*. *Hoot*, Miller's term for aligning the frequency of vocal fold oscillation with every oscillation of the slowest resonance ($1f_o : f_{R1}$), does not just create a fundamental dominant spectrum with a low effective contact quotient.³⁶ *Hoot* is characterized by the purity and strength of resolved harmonics and frequently lacks the buzz of auditory roughness at higher pitches. *Yell*, Miller's term for aligning the frequency of vocal fold oscillation with every other cycle of the slowest resonance ($2f_o : f_{R1}$), is not just the strength of the second harmonic in the spectrum and a higher effective contact quotient.³⁷ It is typically characterized by the brightness and buzziness of the resulting higher-frequency, unresolved energy.

I was particularly fascinated by the *horizontal*, between-vowels, listening implications of Cogan's work—that a pressure pattern in a given frequency range elicits a certain timbral response regardless of the source, and that this is a tool composers and performers can use to generate structure and meaning in music. Every serious voice pedagogy text that addresses acoustics thoroughly explores the idea that different vowels share formants in common.³⁸ For example, I knew from that literature that the vowels [i] and [u] tend to have a similar first formant frequency. I had never seen that concept connected to the idea that those shared formants should sound similar between those vowels. In fact, much of the literature seemed to suggest the opposite: that vowels were the irreducible simultaneous presence of their entire formant structure. What might it mean if a spectral feature common to two vowels produces a similar tone color percept?

Figure 8.2 illustrates the first proof of this concept that I explored. I chose two vowels with a common first formant in a range that allowed me to sing $f_{R1} \approx 1f_0$ on a [u] and an [i]. This aligned a single harmonic with the first resonance in each vowel, making it more likely that the resulting spectral structures would be the same.

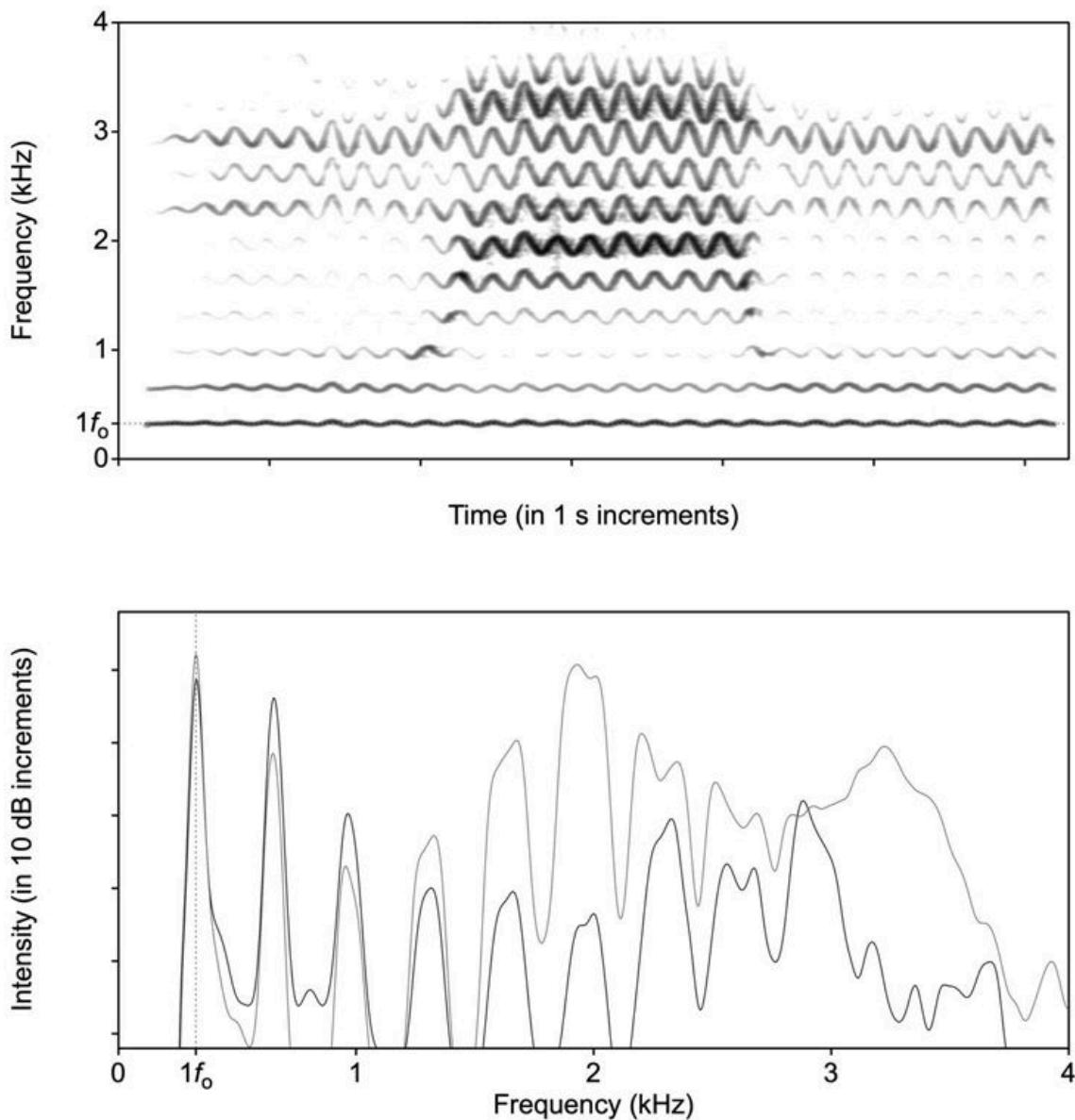


Figure 8.2 A cisgender man singing [u]→[i]→[u] on the pitch E4 (~326 Hz). Spectrogram (top) and spectrum (bottom). A long-term average spectrum for the first [u] is shown in black and [i] in gray. Notice the nearly identical power of $1f_0$ in both vowels. Source: Recorded by author under controlled conditions.

When I pass-filtered $1f_0$, it was impossible to distinguish the [u] from the [i]. There was no tone color transition between the vowels. Repeated blind tests on students and colleagues produced similar results. When asked which vowel the pass-filtered fundamental sounded the most like,

the answer was invariably [u] rather than [i]. When I reintroduced the rest of the spectrum, the [i] could be heard as a separate color rising above an ever-present [u]-like quality. The spectral feature on the screen had a real-world perceptual correlate. It appeared to be both demonstrable and dependable. See Lab #14.

This experience challenges a few long-established ideas. I learned from voice pedagogy texts that vowels are the result of the entire set of formants at once, especially the lowest two. This narrative simplifies the explanation for why two different vowels might share a common formant. If it is the presence of energy at all of those frequencies at once, it would not matter if there were trivial similarities in just a portion of the spectrum. Another way to conceive of this is that vowel quality is equally imprinted across the spectrum. This leads to vowel modification strategies such as passing through a vowel sequence with increasingly faster first formant frequencies (for example, [u] → [o] → [ɔ] → [a] → [æ]) to avoid singing a pitch above the first formant of a given vowel on and above the treble staff.

There is reasonable evidence to support the “entire spectrum at once” hypothesis, as intelligibility of speech is not significantly degraded by filtering only a portion of the spectrum. With some limitations, it does not even matter which portion of the spectrum is filtered.³⁹ This is a prime example of the tension between the intelligibility of speech (based on coarticulation and phonetic context) and the inherent timbre of a sound (based on the response of the ear to spectral and timing information). Though intelligibility may be preserved, the timbre changes radically when the pressure wave is changed in these filtering experiments. The timbre of a sung sound similarly changes when varying the vocal tract shape, registration, loudness, and emotional *affect* while aiming for a consistent phonetic target. See Lab # 15.

The findings of these simple and easily repeated experiments suggest that the first formant of [i] might sound like [u]. That is to say, a strong, simple, vowel-defining formant in [u] becomes a secondary, coloring formant in [i]. Recall Vennard’s (1967) comment that “when one sings Ee ([i]), he is really singing Oo ([u]), plus a still more ringing overtone.”⁴⁰

This experiment suggests that Vennard's framing of vowel formants may be more than a thought experiment. He may have been pointing directly to an actionable aspect of functional listening.

This experiment—a *loose* implementation of the idea of tone color—also provides a cautionary example of how these ideas may be misunderstood. Framed in this way, one might easily counter that an isolated [u] and [i] are each complicated combinations of tone colors. The phoneme [u] contains additional, brighter tone colors and timbral qualities not captured by a pure tone at the pitch E4, even if the tone color of that fundamental is strongly [u]-like. The phoneme [i] is similarly complex and contains both vowel-defining and complementary tone colors at once.

A stricter discussion of the concept of tone color would suggest that [u] and [i] are indeed complexes of varied colors. However, those variations may be parsed out. The *percept* that [u] and [i] share is not itself a *vowel*, even if it finds agreement with [u] more readily than [i]. It is a tone color. It is something that can be noticed, remembered, and anticipated. It is something that can be predicted to have buzziness or not, pitch or not. It is clearly simpler than a full spectrum, but it still elicits specific perceptual qualities that may be contrasted with other parts of each spectrum.

How might this knowledge help voice teachers? Donald G. Miller (2008) suggests the utility of real-time spectrographic feedback because “the teacher’s experienced ear, unaided by objective measurement, has limited ability to determine whether formant tuning is actually taking place.”⁴¹ What is the nature of this limitation? Is it intrinsic to human hearing? Is it not possible to hear how different parts of a spectrum change as pitch and vowel change? Is the issue that we cannot hear what is seen on a spectrogram unaided? Or is the issue that the spectrogram displays information in a manner that does not readily reflect how a human perceives complex sounds?

I argue that the ability to aurally parse out the consistent color of that shared first formant (the fundamental here) in both the [u] and [i] vowels offers an introduction to how we may ultimately resolve this limitation. Is a singer achieving a $f_o : f_{R1}$ coupling on an [i] vowel? If the teacher can

parse out the perceptual quality of $1f_o$ (warm, pure, [u]-like in this case) from the rest of the spectrum, they may literally listen for it in real time.

OPEN QUESTIONS AND POTENTIAL APPLICATIONS

The above material leaves me with an unanswered question: What is the underlying physiological or cognitive mechanism for tone color perception? At this moment I am honestly not sure, but I can speculate. Perhaps the basilar membrane transmits information in qualitatively different ways from different tonotopic areas. Perhaps tone color is a cognitive response to the timing of nerve firings sent to the brain. Perhaps it is the result of deeper cognitive processing that the auditory cortex layers onto its analysis of the incoming signals. Like the basilar membrane, the auditory cortex itself may be organized tonotopically.⁴²

Perhaps more interesting questions include: Why do some listeners notice the phenomenon of tone color without any training while others do not? Multiple people have noted and published about this over the past 140 years. One assumes that this perceptual property of simple sounds has always occurred to *some* people. How can we not notice that the *coo* of a mourning dove is dissimilar from a high-pitched whistle? Similarly, why does this selective hearing skill appear to be so easily teachable? How does the acquisition of this critical listening skill change a listener's percept so rapidly? How do these perceptual changes persist as cognitive processes over time?

As with auditory roughness and pitch perception, discussed in [chapter 7](#), the idea of tone color maps well to a simple, two-timbre understanding of the sound of a singer. If you would expect a given sound to show strong lower frequency aspects in a spectrum and lack much higher frequency energy, you will favor the tone color palette that sounds more like [u] than [i]. Similarly, any sound that is extremely bright (regardless of the target vowel) will have an additional tone color component that sounds more like [i] than [u]. We will investigate these ideas in more detail in [chapter 10](#).

DISCUSSION QUESTIONS

- What is a simple definition of *tone color* as discussed in this chapter? If you were to make that definition more complex, what additional information would you add?
- If a formant is a strong spectral peak (meaning it is a feature of a spectrum), does it represent persistent frequency information at all timescales? If not, by what process is the persistent formant generated from a dynamic waveform?
- Characterize the difference between frequency and brightness.
- How can you reconcile the idea that a range of formant frequencies can convey a given vowel with the idea that changes in those formant frequencies change tone color? Is tone color the same thing as a vowel?
- What is the earliest reference in this chapter to the idea that low- and high-frequency pure tones elicit different timbres?
- What might be meant by the two-part statement: Tone colors are not vowel-like; vowels are tone color-like?
- Have you encountered the idea of tone color in the acoustics, perception, or cognition chapters in other voice pedagogy or vocology texts? What might account for the relative presence or absence of this idea compared to other aspects of acoustics, perception, or cognition?
- When you carried out Lab #14 and heard the [u]-like quality of the first formant of the [i] vowel, what implications might there be for the way you cue the [i] vowel in a singer? Based on what you now know, is [i] a bright vowel or a warm vowel?
- What has been your experience using a spectrogram for biofeedback while singing? Considering the ideas of auditory roughness and tone color already covered, do you think you saw what you heard?

NOTES

1. Ingo R. Titze et al., "Toward a Consensus on Symbolic Notation of Harmonics, Resonances, and Formants in Vocalization," in *Journal of the Acoustical Society of America* 137, (2015): 3005–7.
2. Ludimar Hermann, "Beträge zur Lehre von der Klangwahrnehmung," in *Pflueger's Archiv*, 56 (1894): 467–99.
3. C. Julian Chen, *Elements of Human Voice* (New Jersey: World Scientific, 2017), ix.
4. Reinier Plomp, "Experiments on Tone Perception" (PhD diss., Institute for Perception RVO-TNO, 1966), 131–32.
5. Barbara M. Doscher, *The Functional Unity of the Singing Voice*, 2nd ed. (Lanham, MD: Scarecrow Press, 1994), 134–35.
6. William Vennard, *Singing: The Mechanism and the Technique* (New York: Carl Fischer, 1967), 130.
7. Andrey Anikin, Katarzyna Pisanski, and David Reby, "Static and Dynamic Formant Scaling Conveys Body Size and Aggression," in *Royal Society Open Science* 9: 211496, <https://doi.org/10.1098/rsos.211496>.
8. Ingo R Titze, "Human Speech: A Restricted Use of the Mammalian Larynx," in *Journal of Voice* 31, no. 2 (March 2017): 135.
9. Mark H. Ashcraft, *Cognition* (Upper Saddle River, NJ: Pearson Prentice Hall, 2006), 382–83. Ashcraft features a discussion of coarticulation. Terrance M. Nearey, "Static, Dynamic, and Relational Properties in Vowel Perception," in *Journal of the Acoustical Society of America* 85, no. 5 (May 1989): 2088–113. Nearey offers a discussion of conflicts between the context effect and inherent quality in speech research.
10. Peter B. Denes and Elliot N. Pinson, *The Speech Chain: The Physics and Biology of Spoken Language*, 2nd ed. (New York: W. H. Freeman, 1993, reissued 2015), 143.
11. Kenneth Bozeman, *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013); Berton Coffin, *Coffin's Sounds of Singing: Principles and Applications of Vocal Techniques with Chromatic Vowel Chart*, 2nd ed. (Lanham, MD: Scarecrow Press, 2002).
12. Ernst Mach, "Zur Analyse des Tonempfindungen," in *Sitzungsbericht der kaiserlichen Akademie der Wissenschaften in Wien*, Bd. 92, Abt. 2 (1885): 1285.
13. Bozeman, *Practical Vocal Acoustics*, 14–15.
14. Kenneth Bozeman, *Kinesthetic Voice Pedagogy 2: Motivating Acoustic Efficiency*, expanded edition (Delaware, OH: Inside View Press, 2021), 54–56, 147.
15. Ian Howell, "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color" (DMA diss., New England Conservatory of Music, 2016), 47.
16. Mach, "Zur Analyse der Tonempfindungen," 1283–89.

17. Gustav Engel, "Ueber den Begriff der Klangfarbe," in *Philosophische Vorträge* (Berlin), Neue Folge, II. Ser., Heft 12 (1886): 311–55.
18. Carl Stumpf, *Tonpsychologie*, vol. 2 (Leipzig: S. Hirzel Verlag, 1890), 514–43.
19. Wolfgang Köhler, "Akustische Untersuchungen I," in *Beiträge zur Akustik und Musikwissenschaft* 4 (1909): 134–82, reprinted in *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 54 (1910): 241–89; Wolfgang Köhler, "Akustische Untersuchungen II," in *Zeitschrift für Psychologie* 58 (1911): 59–140, reprinted in *Beiträge zur Akustik und Musikwissenschaft* 6 (1911): 1–82; Wolfgang Köhler, "Akustische Untersuchungen III," in *Zeitschrift für Psychologie* 72 (1915): 1–192.
20. Hermann Grassmann, "Ueber die physikalische Natur der Sprachlaute," in *Annalen der Physik und Chemie* 1, (1877): 606–29.
21. K. von Wesendonk, "Ueber die Synthese der Vokale aus einfachen Tönen und die Theorien von Helmholtz und Grassmann," in *Physikalische Zeitschrift* (1909): 313–16.
22. A. P. Weiss, "The Vowel Character of Fork Tones," in *The American Journal of Psychology* 31, no. 2 (April 1920): 166–93.
23. V. Engelhardt and E. Gehrcke, "Ueber die Vokal-charaktere einfacher Töne," in *Zeitschrift für Psychologie* 115, (91930): 16–33.
24. John Morton and Alan Carpenter, "Judgement of the Vowel Colour of Natural and Artificial Sounds," in *Language and Speech* 5, no. 4 (1962): 190–204.
25. Plomp, "Experiments," 131–32.
26. William Vennard, *Singing: The Mechanism and the Technique* (New York: Carl Fischer, 1967), 128.
27. Reinier Plomp, *The Intelligent Ear: On the Nature of Sound Perception* (London: Lawrence Erlbaum, 2002), 23–24.
28. Fritz Winckel, *Music, Sound and Sensation: A Modern Exposition* (New York: Dover, 1967), 24.
29. Wayne Slawson, *Sound Color* (Berkeley: University of California Press, 1985), 20.
30. Robert Cogan, *Music Seen, Music Heard: a Picture Book of Musical Design* (Cambridge: Publication Contact International, 1998), 110.
31. Julian Anderson, "Spectral music," Grove Music Online (2001); accessed 4 August 2021.
32. Julian Rushton, "Klangfarbenmelodie," Grove Music Online (2001); accessed 4 August 2021.
33. Robert Cogan, *New Images of Musical Sound* (Cambridge, MA: Harvard University Press, 1984), 7 and 12; Cogan, *Music Seen*, 110; Winckel, *Music, Sound and Sensation*, 14.
34. Donald G. Miller, *Resonance in Singing: Voice Building through Acoustic Feedback* (Princeton, NJ: Inside View Press, 2008), 68.

35. Miller, *Resonance in Singing*, 74.
36. Miller, *Resonance in Singing*, 52.
37. Miller, *Resonance in Singing*, 54.
38. Ian Howell, "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color" (DMA diss., New England Conservatory of Music, 2016), 60. This dissertation offers a selection of images from: Scott McCoy, *Your Voice: An Inside View*, 3rd ed. (Delaware, OH: Inside View Press, 2019), 68–70, 72; Kenneth Bozeman, *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013), 13; Barbara M. Doscher, *The Functional Unity of the Singing Voice*, 2nd ed. (Lanham, MD: Scarecrow Press, 1994), 138, 152; Sundberg, *The Science*, 107; Peter B. Denes and Elliot N. Pinson, *The Speech Chain: The Physics and Biology of Spoken Language*, 2nd ed. (New York: W. H. Freeman, 1993, reissued 2015), 143.
39. Denes and Pinson, *The Speech Chain*, 140–83. This text offers a thorough description of these paradoxes.
40. Vennard, *Singing*, 130.
41. Miller, *Resonance in Singing*, 21.
42. Colin Humphries, Einat Liebenthal, and Jeffrey R Binder, "Tonotopic Organization of Human Auditory Cortex," in *NeuroImage* 50, no. 3 (2010): 1202–11, doi:10.1016/j.neuroimage.2010.01.046.

9

Absolute Spectral Tone Color

The exploration of tone color in the previous chapter suggests that the study of sung sounds may be problematized by a framework that prioritizes language cognition over inherent timbral percepts. Singers *do* frequently leverage language to communicate, and much related to musical phrasing draws on the cadence and prosody of text. However, the timbral variations potentially employed by a singer far exceed the inventory of sounds necessary for speech intelligibility. A singer may substitute more strongly resonant vowels for problematic ones, leveraging the listener's cognition to preserve intelligibility despite the actual sounds presented. The singer may subtly adjust a vowel to create a favorable acoustical environment in the vocal tract for a specific registration transition or "money note." Additionally, a framework that is based on language cognition lacks a mechanism to understand timbral changes related to the expanded pitch and dynamic ranges that are common in singing, yet largely absent in everyday speech. Entire genres of music feature singing characterized by timbral qualities that are either irrelevant to the intelligibility of language or purposely obfuscate intelligibility to those outside a subculture.

As voice educators who engage a singer's sound on a technical level, we are best served by discussing the actual sounds being made. This is true even though such sounds also convey linguistic and paralinguistic meaning. Therefore, it seems reasonable and necessary to consider

models for understanding the sound of a voice through a language-agnostic lens.

As explored in [chapter 8](#), it is not controversial to suggest that two pure tones of the same frequency are equally *bright*, and that two pure tones of differing frequencies differ in *brightness*. One may also extend, with some important caveats, the tone colors of pure tones to describe aspects of a complex wave with a similar dominant frequency band. The tone colors of pure tones then serve as references for understanding the ongoing spectral flux of tone colors in a complex sound wave. This is the primary practical application of what I have dubbed *absolute spectral tone color*.

ABSOLUTE SPECTRAL TONE COLOR: HISTORY AND DEFINITION

Before diving into the definition of absolute spectral tone color, I want to share how I arrived at this term, and why these specific words matter. In 2013 I offered the term *absolute timbre* to point to qualitative aspects of the tone color qualities explored in chapter 8.¹ As Cornelia Fales (2002) quips, “Timbre is a slippery concept and a slippery percept, perceptually malleable and difficult to define in precisely arranged units.”² Yet Plomp supports the idea that “simple tones have timbre” related to their objectively measurable frequency—in other words, their brightness.³ I saw the term *absolute timbre* as a way to point to that narrow tone color-related aspect of timbre that appears to have an objective quality without suggesting that timbre in its entirety was objective. If tone color relates to frequency and only applies to pure tones, there are no further timbral considerations to muddy the tone color percept of that pure tone. However, it became quickly apparent that the term *timbre* used in this way problematically denotes other qualities of a sound beyond the tone color itself, regardless of the context provided.

In place of *absolute timbre*, I started to use the term *absolute spectral tone color* (ASTC). This new term calls attention to the attribute of timbre in question: It is the brightness itself, rather than how the presence of brightness fluctuates over time. The term *spectral* points to the relationship of the frequency domain to brightness as a timbral property. The unqualified nature of the word *absolute*—a fighting word when discussing perception—honors the idea that for a simple tone, the frequency determines both pitch and timbre. Thus, two pure tones of the same frequency, separately produced, will share a common tone color percept.

A short sidebar: As loudness is controlled for in the ANSI definition of *timbre*, the pitch percept—and therefore the tone color—of a pure tone varies somewhat with intensity. In this limited case, the way the human

ear processes the frequency of a pure tone stimulus affects not just pitch perception but also our perception of the frequency.

You can explore this idea by doing the following: Generate a sine wave at 100 Hz using Audacity, Praat, Madde, VoceVista Video Pro, or another compatible program of your choice. Use headphones, which will nullify potential head position-related phasing issues when listening to a pure tone over speakers. Slowly decrease the playback volume from loud to barely perceptible. You will notice that the pitch percept rises slightly as the sound becomes quieter. Repeat this experiment with a 3 kHz tone and you will notice that the opposite occurs: As you decrease the playback volume, the pitch percept lowers slightly.⁴ The degree of intensity-modulated pitch drift is not so severe that it noticeably changes the tone color of the sound; but, strictly speaking, the measurable frequency of a pure tone is not always the absolute correlate to the pitch percept. See Lab #16.

Most periodic sounds we encounter are at least a little bit complex, and the pitch percept variability of pure tones will likely never rise to the level of your attention in the voice studio. However, this does help to explain a phenomenon that always bothered me and that you may have also noticed. If I listen to music that begins with several bars of a dull-sounding bass guitar and no other pitched instrument, I am frequently surprised by the key when chordal or melody instruments join. Once those more complex patterns vibrate along with the simpler bass sound, that bass sound snaps into the right key in my perception. Those kinds of bass sounds are too close to a pure tone to trigger the correct pitch percept on their own. With this sidebar concluded, I will proceed with the definition of ASTC.

In my 2016 dissertation I define *absolute spectral tone color* in the following way:

Any two or more simple sounds (for example a sine wave, [a pass-filtered] harmonic of a complex tone, or narrowly notch filtered band of noise) of identical frequency, regardless of their sources, will produce an identical tone color percept independent

of other spectral fluctuations considered aspects of timbre. If these simple sounds are located within a complex sound, their inherent absolute spectral tone color is never lost or changed, only expressed or masked. These tone colors may be placed on a continuum and bear a meaningful similarity to several vowels.⁵

If one describes the tone color of simple sounds along a continuum of brightness (as Plomp [1966] suggests is long accepted in psychoacoustics),⁶ the principle of absolute spectral tone color requires that we label any two simple sounds of the same frequency with an identical brightness value. Similarly, if we adopt Cogan's [(1998)] register-based system and divide timbre into grave, neutral, and acute regions, two simple sounds of the same frequency exhibit exactly the same quality of graveness, neutrality, or acuteness.⁷ It matters less that we use a specific scale [or labels] than that we make the conceptual leap that there are absolute values along the continuum.⁸

The underlying logic here is that recognizing the similarity in tone color of two separately produced pure tones at the same frequency (and intensity to a very minor extent) is more important than how one confronts and labels that phenomenon. This is the central idea of ASTC, which acknowledges how limited this perceptual property is. Figure 9.1 illustrates this scale with labels I assigned to help us understand the dominant spectral signifiers of certain vowels. These labels also serve as a way to identify and understand the tone color contributions of complementary, *non-vowel-defining* portions of any spectrum. This includes energy below F_1 and above F_2 and demonstrates that only one vowel formant (frequently but not exclusively F_2) provides the identifying tone color of the vowel. The methodology for deriving these boundaries is explored in greater detail in Howell (2016). My earlier findings are based on listening experiments using pure tones and small clusters of harmonics.⁹ If you are interested in exploring this, please look at Lab #17.

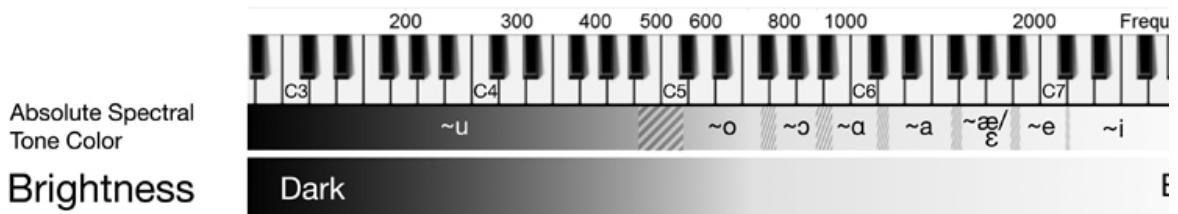


Figure 9.1 Author's scale of absolute spectral tone color. Source: Howell, "Parsing," 73; Howell, "Necessary roughness," 5.¹⁰

ABSOLUTE SPECTRAL TONE COLOR: LIMITATIONS

Why use vowel symbols as labels? This is a question I have struggled with. Remember that the tone colors of pure tones have historically been challenging to label outside the use of vowel names. Vowels are how humans appear to think about tone color. Ask a child to imitate the *coo* of a mourning dove and they will intuitively make a lip-rounded [u] shape. Ask them to imitate the bleat of a sheep and they will shape laterally for [æ]. Ask them to make the squealing sound of wet city bus brakes and they will screech a thin [i] sound. We imitate the sound of an ambulance siren with *wheeeooowheeeoo* [wijuwijuwiju]. We all seem to understand on a deep level that certain vowels, as objective sounds outside of their semantic function in language, are characterized by certain tone colors. We all seem to use vowels to generate spectra that imitate sounds that share the same tone colors, despite not being generated by a human voice.

Even yet, to use a vowel label in this work, even when qualified with softening language such as *vowel-like*, is to invite critique. It is easy to assume that those vowel labels imply that a simple sound characterizes the complete timbral quality of the vowel percept it references. This argument has been ongoing for more than a century. Stepping to the side of such arguments, I want to qualify ASTC. First, it is important to reiterate that the tone colors of pure tones are only *similar* to the *defining quality* of the chosen vowel; they are vowel-like. Actual vowels are complex combinations of many different tone colors *and other time-variant and time-invariant aspects of timbre*. It should be obvious to anyone that a pure tone will never fully represent that sound. Second, although the separate ASTC regions labeled in figure 9.1 indicate frequency regions of similar vowel-like tone colors, even the tone colors within each range lie on a continuous spectrum. The vowel-like tone color on the border between two ranges will share an affinity with both, much like the transitions between dominant colors in a rainbow.

The simplest notation for an ASTC label is to place a tilde (~) before the corresponding IPA symbol. For example, the ASTC of a simple sound with a frequency of the pitch D4 is ~u. The tilde conveys *like* or *approximately* the defining tone color of [u]. More formally, these ASTC symbols may then be set apart from surrounding text with *pipes* (also called *vertical bars* or *vertical lines*). This follows the convention of the International Phonetic Alphabet, which uses punctuation (usually solidi or square brackets) to parenthesize its symbols. Visually, these pipes look identical to the absolute-value brackets used in mathematics, which aligns with the concept. The tone color of a pure tone at the pitch D4 may then be notated as |~u|.

If you are aware of the order of second formant (F_2) frequencies for the vowel labels I use, you will notice that the chart in figure 9.1 follows that pattern with the general exception of |~u| and |~o|. I write “general” because one could certainly close and round a [u] or an [o] to the point that both formants fell below the average values for that vowel. But in many cases, these vowels are strongly defined by the tone color of the first formant (F_1), while the second formant (F_2) provides a weaker, brighter complementary tone color. To someone committed to the notion that the entire formant structure defines the entire vowel, this may seem like a suspiciously unsupported parsing of the spectrum. However, it is easily demonstrated that the second formant (F_2) is the most highly variable formant among vowels (see [figure 9.2](#)). If you think about formant contributions in these terms, it should at least pique your curiosity as to whether the second formant (F_2) carries much of the burden of vowel identity. Or, put another way, how much of the *vowel-defining tone color* of the spectrum is strongly characterized by the second formant? Removing that second formant may not obliterate intelligibility, given the integral role of coarticulation in language cognition. However, removing F_2 definitely changes the timbre in a more radical way relative to removing any other formant (with the likely aforementioned exceptions of [u] and [o]). See Lab #18.

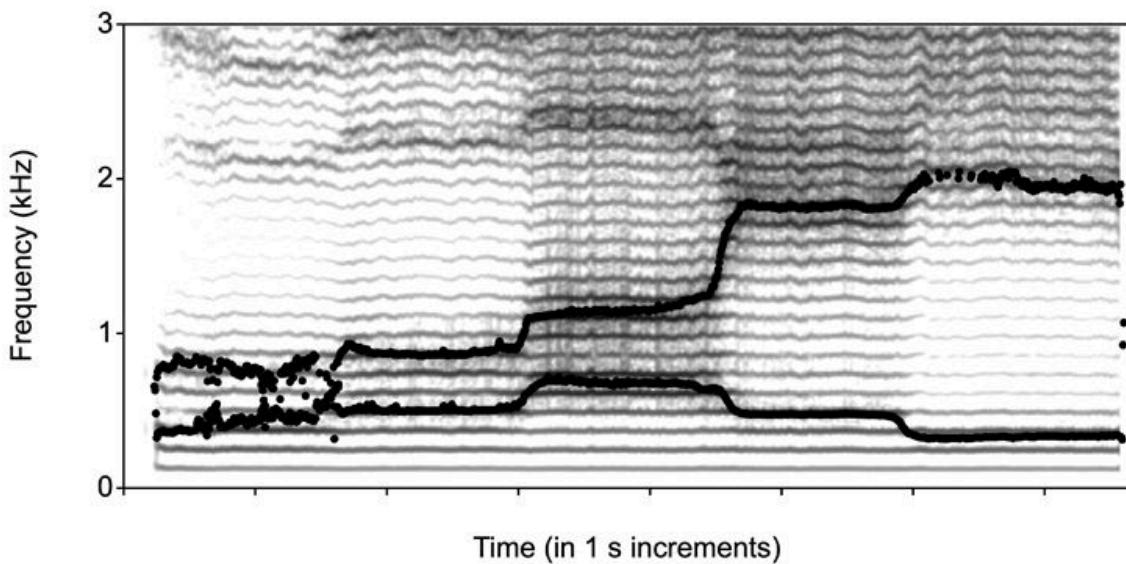


Figure 9.2 Spectrogram of [u], [o], [a], [e], [i] showing the first and second formants. Source: Recorded by author under controlled conditions.

I bring two obvious biases to this schema: (1) I limit the frequency range worth labeling to reflect the potential range of vowel formants in the human voice. Acknowledging our vocologist and voice science colleagues who study nonhuman subjects, I am sure that tone color varies in meaningful ways below (slower than) |~u| and above (faster than) |~bright i| for other animals. I suspect that bats, mice, rats, and other animals capable of very high-frequency communication may perceive differences in a range that a human would never hear, let alone differentiate from the brightest |~i| imaginable. If we could produce a strong vowel formant lower than the F_1 of [u], I can imagine an inventory of additional, darker vowels experienced by some other creature. Therefore, the ASTC scale is strongly anthropomorphized.

(2) While the sounds associated with specific IPA labels do vary between languages, I have limited these labels to a subset of familiar vowel sounds that occur in English. I would expect non-English speakers to subtly prioritize other labels where they are spectrally relevant. For example, a native German or French speaker may argue that |~y| has a place in or around where an English speaker finds the transition between

$|\sim e|$ and $|\sim i|$, because they are familiar with the strong spectral energy that exists in [y] in the space between and around the second major spectral peaks of [e] and [i]. ASTC holds, regardless of how one confronts the tone color differences between these peaks, because all that matters is that one would consistently use the same label when comparing a pure tone to a novel pure tone of the same frequency. Even if a non-English speaker would argue for slightly different borders for these ASTC symbols, I am confident that they would do so consistently.¹¹

The ASTC framework then *does not depend on the agreement* between two individuals that a given label is perfect. Instead, it invites these individuals to acknowledge that two similarly presented pure tones of the same frequency dependably elicit the same tone color percept. And that regardless of how they would assign the label, they would dependably do so. Think in terms of dark and bright if you wish—or grave, neutral, and acute, in a nod to Cogan. ASTC allows for a nuanced exploration of a singer’s sound, but less granular approaches still have value, especially in the voice studio.

One final point regarding the ASTC labels in figure 9.1: All of these labels reference vowels that were selected for their typically strong spectral peaks in the related tone color region. So we find [a], but not [Λ]; likewise, [i] but not [I]; [ε] as paired with [e] appears to be an exception to me. Especially in the context of considering these pairs to represent the clear (tense) versus neutralized (lax or centralized) versions of a near-similar sound, we can try to make a connection to the acoustical model from [chapter 6](#). In a spectrum of a sustained, speech-pitch sound, we would find the transition from [a] to [Λ] characterized by lowering F_1 . In the example shown in figure 9.3 (top), note the strength of the F_1 for [a] with a clearly defined peak at $4f_o$ and strong surrounding harmonics. If you imagine the f_{R1} resonance—which acts as a potential amplifier in the sense of the source-filter model—moving to the *left* in figure 9.3 (bottom) for the [Λ], you can see that the F_1 peak has flattened out. The energy surrounding the second formant (F_2) peak between $7f_o$ and $8f_o$ also decreases in intensity.

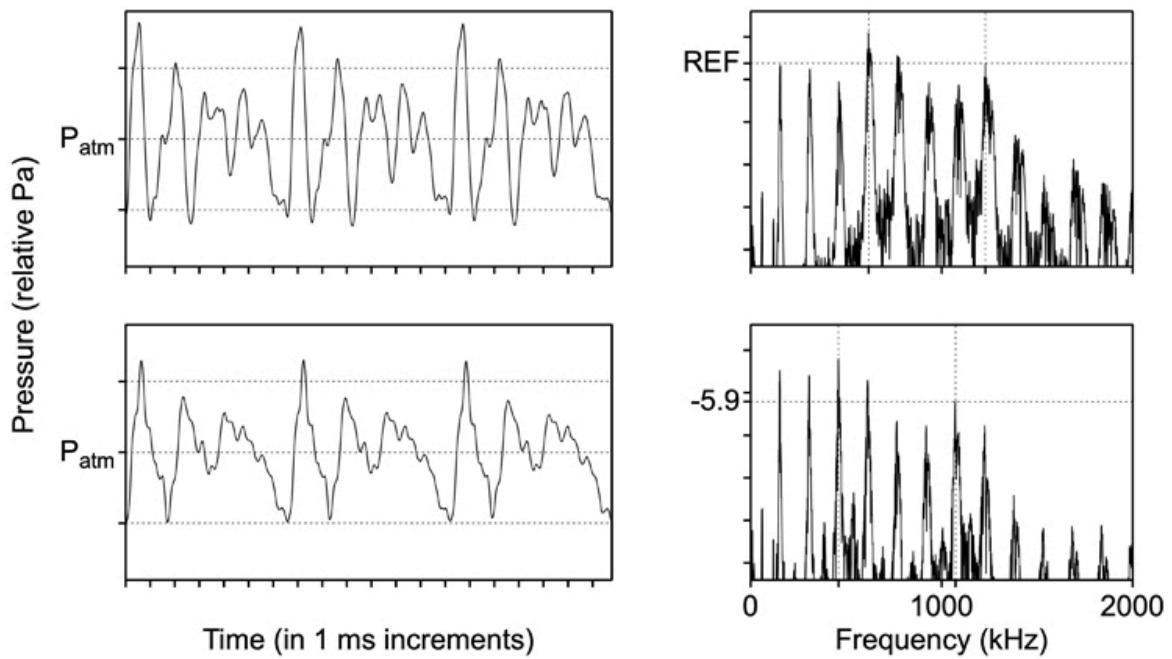


Figure 9.3 Waveforms (left) and spectra (right) of a cisgender male sustaining the vowels [a] (top) and [ʌ] (bottom). Source: Recorded by author under controlled conditions.

A look at the waveforms for these two samples brings another dimension to our inquiry. Figure 9.3 shows just three periods of the [a] (top) and [ʌ] (bottom) waveforms. These samples were selected from high-amplitude portions of each waveform, as ongoing amplitude modulation (shimmer) is a feature of human phonation. The waveform for the [a] is not only higher in amplitude overall but remains higher in amplitude through each pitch period. Here we see the difference in the vocal tract damping of these two vowels. The [a] resonates with greater strength and transfers energy more efficiently from pitch period to pitch period. In the spectra, note that the F_2 peak is ~ 6 dB stronger in [a] and that the F_1 both lowers in frequency and flattens out as a spectral feature in [ʌ]. These spectral differences have correlates in the decreased amplitude of the pattern of pressure in the air. If you pass-filter the formants around the third to fourth ($3f_o$ – $4f_o$) and seventh to eighth harmonics ($7f_o$ – $8f_o$) (the spectral representations of the strongest

oscillations within the repeating pressure pattern) for these two vowels, you will hear that their ASTC does not change much. Those tone colors are just expressed more strongly (loudly) in the [a].

Building on the [u] and [i] example in [chapter 8](#) (Figure 8.2), consider how moving from [u] to [i] to [æ] on the same pitch might impact the power of the fundamental, when f_{R1} aligns with the fundamental of [i] and [u], but not [ae] (see figure [9.4](#)). Note that the fundamental in [i] is ~11.7 dB stronger than in [æ], and ~2.3 dB is stronger than in [u]. In all three instances, the tone color contribution of that energy is the same. In my ASTC framework (see figure 9.1), this energy at D4 (~294 Hz) sounds [u]-like: |~u|. This color is strong within the [i] and [u]. It is also present in the [æ], but weaker. These vowels do not just differ linguistically; their tone color content differs as well. The defining characteristic tone color of [u] becomes a strong secondary color in [i], and relegates to a background color in [æ]. That tone color is either strong or it is not. The tone color of energy at that frequency does not change between these vowels. See Lab #19.

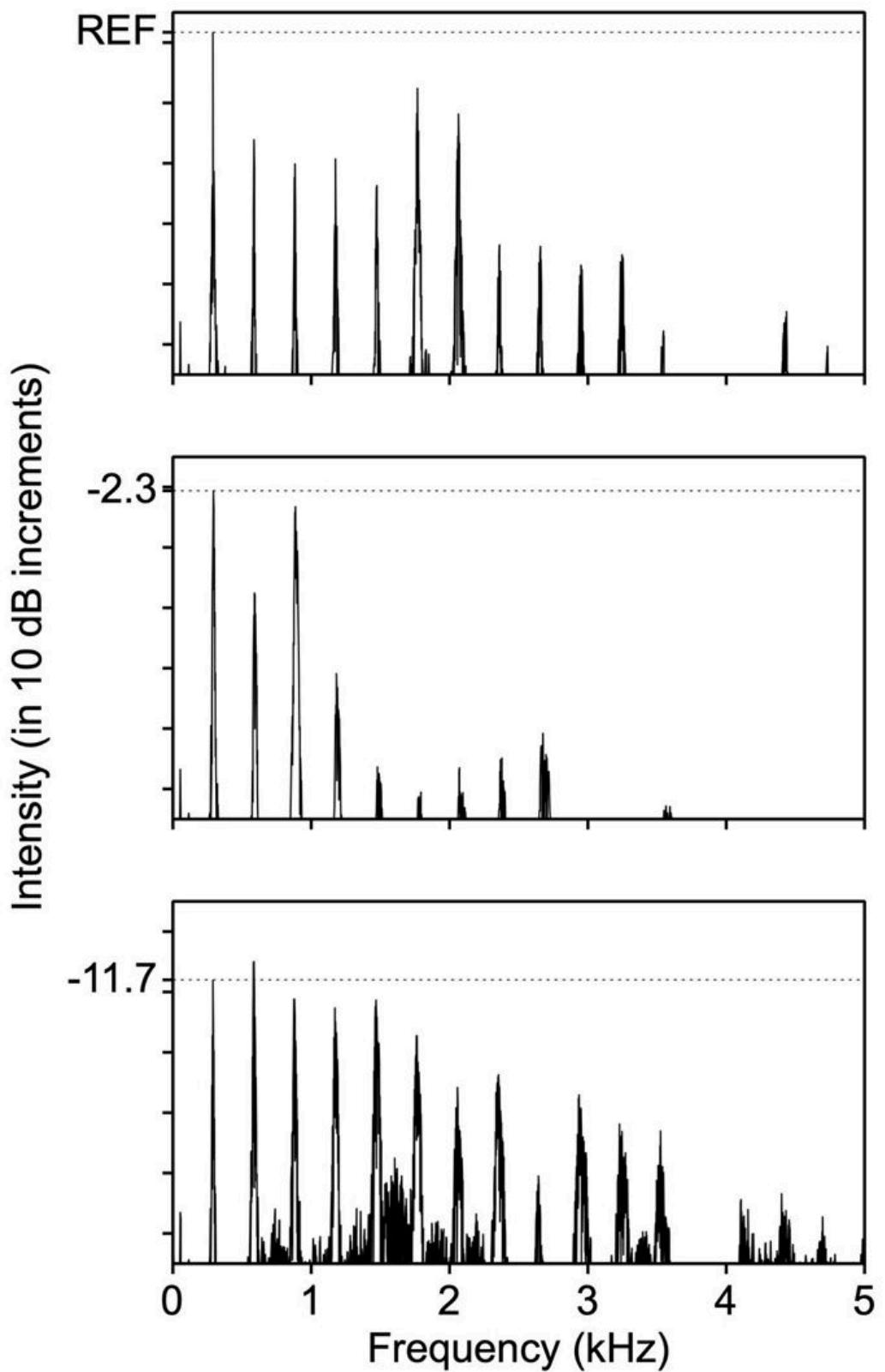


Figure 9.4 A cisgender man singing [i] (top), [u] (middle), and [æ] (bottom) on D4. Note the dramatic difference between the levels of f_0 between the [æ] and the other two vowels. Source: Recorded by author under controlled conditions.

If you know how to *listen for* that $|~u|$ tone color in all three vowels, you can then evaluate its relative strength without a computer. If the $|~u|$ tone color is suppressed in the context of an [i] vowel at this pitch, the resulting sound will likely be thin. You may prize that if you sing into a microphone that automatically boosts the low end of the spectrum. If the $|~u|$ tone color is stronger than usual in the [æ], the resulting sound will likely be warmer than usual. You may prize that if you sing unamplified music on the treble staff and want to generate a high-amplitude pressure wave. All of these choices are fine aesthetically, but these vowels have very different tone color baselines. The point of ASTC in this context is to be able to notice and effectively describe these baselines and technically relevant deviations.

EXPLORING ABSOLUTE SPECTRAL TONE COLOR EXPERIENTIALLY

The basic idea of ASTC is challenging to study scientifically. Play a series of pure tones to naive listeners and you will invariably get a variety of responses to the question “What vowel is this like?” This is to be expected, as few listeners have the grounding to understand this question in terms of the ASTC framework. As we have seen, vowels are complicated combinations of tone colors, buzzy and pure qualities, and resolved and unresolved pitch percepts. The question itself is imprecisely put and does nothing to articulate the specific aspect of timbre captured in ASTC. To explore ASTC with a new listener, again reference Lab # 15.

VOWELS SHARE COMMON TONE COLORS

In previous chapters I have mentioned the idea that vowels share common formants. [Figure 9.5](#) is a scatterplot graph of the first (F_1) and second (F_2) formant values for adolescent cisgender male speakers after Weenink (1985), which hints at these patterns.¹² Any two vowels that share a position on the horizontal axis share a similar first formant (F_1) frequency. As I mentioned in [chapter 8](#), the wide range of values captured within these ellipses do not suggest that there is no timbral difference between the vowel sounds within the ellipses. Instead, this graph suggests that these ranges of formant values were demonstrated to successfully fulfill their semantic roles in the context of speech.

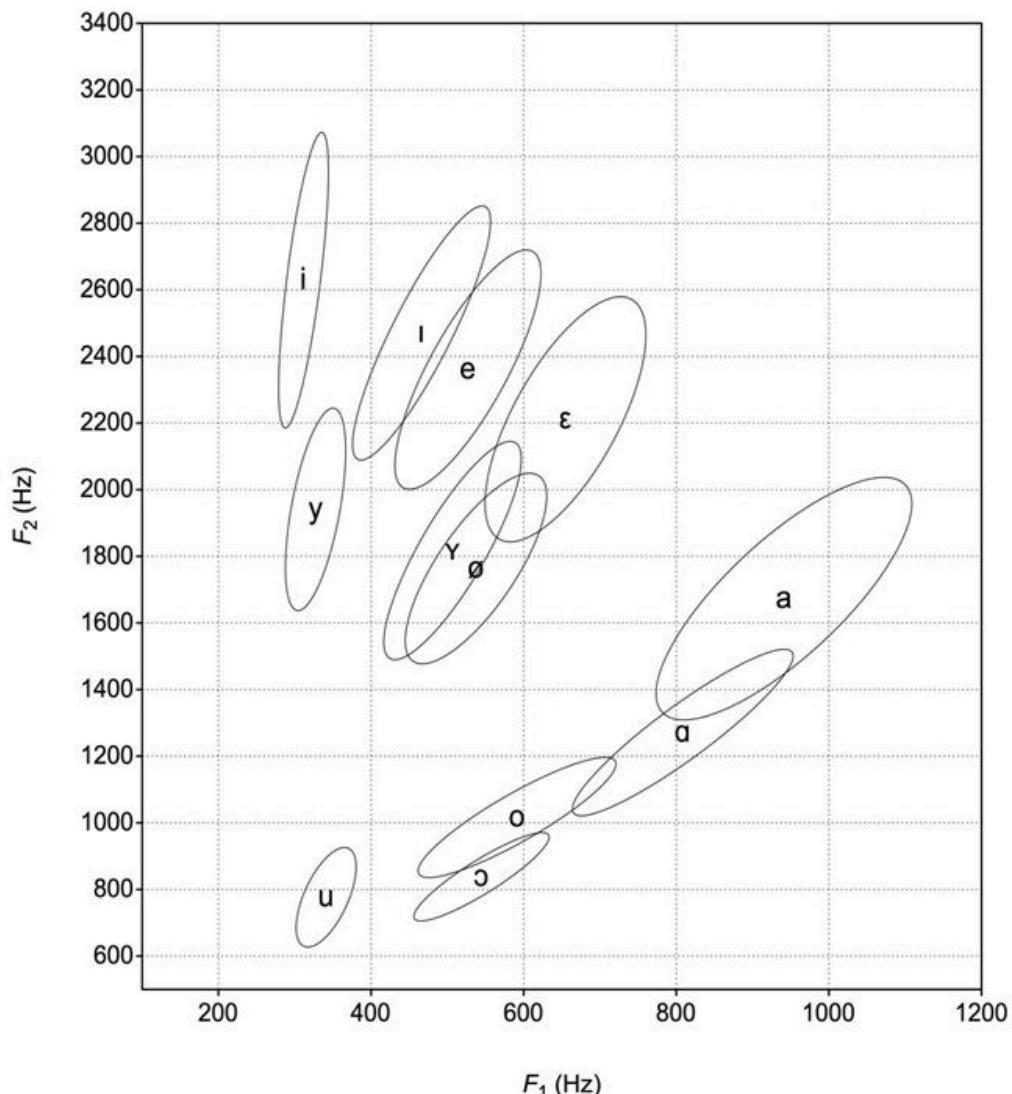


Figure 9.5 Scatterplot graph of first (F_1) and second (F_2) formant values for adolescent cisgender male speakers after Weenink (1985). Source: Generated by author in Praat.

Locate [u] in the lower-left corner. Moving up in the graph you will find that [y] and [i] align with the horizontal values for [u]. We may not intuitively think of [y] and [i] as *warm* vowels, but our explorations demonstrate that those vowels potentially share an |~u| warmth similarly featured in an [u]. Moving up the chart from [o] also points to potential

shared first formants (F_1) with [e] and other vowels. If you check the ASTC range of those first formant (F_1) values, it will have opened from |~u| to |~o|. This is also sensitive to the pitch being sung. If these vowels are modified, either by changes in the vocal tract shape or as pitch-related timbral shifts, those tone colors may shift as well. For example, one may sing a credible [i] well above the first formant value (F_1) for that vowel, so long as the second formant (F_2) remains in the |~i| tone color range.

If you would like to create your own chart like this, perhaps sourced from your own students in a classroom setting, see Lab #19.

POTENTIAL CONCLUSIONS AND APPLICATIONS

The essential takeaway from this chapter builds on the ideas from [chapter 8](#): We can (1) notice tone color as a property of pure tones, and (2) notice that the tone colors of different frequency pure tones are experientially different and similar frequency pure tones are experientially similar. In this chapter I have proposed what I think is a useful and logically ordered series of vowel labels to represent the ranges of the frequency spectrum that correspond to the human voice. Further, I explored that anthropomorphizing these different tone colors in this manner is both useful and problematic.

This anthropomorphizing is useful because basic ideas like open and close¹³ vowels have identifiable timbral correlates. For example, [a] is not just linguistically more *open* in speech (with the tongue dorsum further from the roof of the mouth) than [o]; we can also hear the vowel formants shifting to brighter tone colors. At the same time, using vowel labels from speech is problematic because spoken and sung vowels elicit more complex percepts than pure tones do. These labels in the ASTC schema risk conflating the two. Additionally, it is problematic because ASTC applies to all sounds we hear, not just human vocalization. Recall that the call of a mourning dove does not sound like the defining quality of a bright [i], nor does the squeal of wet bus brakes resemble the dominant tone colors of a [u]. ASTC can inform both spoken and sung sounds, yes. But ASTC is not exclusively related to language cognition.

One of the delightful implications of the above information is that vowels are assemblages of tone colors. Yes, one may comprehend speech between speakers despite differing formant frequencies, and in the speech-pitch range, the spectrum is in continuous flux. The entire venture accommodates variability. However, encountering a vowel as a musical sound allows one to contemplate several ideas relevant to singers and voice teachers. It suggests that given a particular pitch and intensity, attending to specific tone color aspects of the sound allows us to fine-tune a vowel. It also primes us for the exploration of commonalities

between different sounds. The commonalities point to the idea that different functional coordinations of the singing voice may share qualities in common. This suggests pathways to realize *adjacent* or *intermediate* functions that might help a singer to discover the feeling or sound of the actual target function. More on this in [chapters 11](#) and 12.

In the following chapter I will explore how ASTC exists in “the wild” as part of the messy and ever-changing percept of a complete vowel, and I will explore its intersection with pitch perception and auditory roughness. I continue to encourage the reader to think of ASTC as *informing* rather than dictating the resulting percept.

DISCUSSION QUESTIONS

- Explore saying the phrase “My, it is hot today” in multiple different accents, character voices, or with different emotional affects. Be inventive. Really try to distort the vowels. Can other people still understand what you said? Record that sound in these different ways and excise only the “o” in *ho t*. How different do these vowels sound in isolation? What do you think this suggests about the difference between the inherent timbres of each of those sounds and the process of language cognition?
- Do the ongoing spectral fluctuations of speech mean that perception of brightness is inherently subjective? Can you think of an example of a natural or mechanical sound that *sounds* like the defining aspect of a specific human vowel?
- Do we ever experience actual sine waves (perfectly symmetrical, continuous changes in pressure) in nature? What are some sounds that are close to sine waves?
- Think about the complexity of a vowel relative to the specificity of the ASTC of a pure tone. What do you think about the hierarchical idea that vowels may have certain spectral features that *define* the vowel and other spectral features that are *complementary* tone colors? Can you think of any implications for this idea given the wide pitch and intensity range used by singers?
- What are some arguments against using IPA vowel symbols for ASTC? What are some arguments in favor?
- Does the principle of absolute spectral tone color require that two people agree about a specific label? What does it require?
- If vowels share formants, give a few examples of potential timbral similarities between different vowels.

NOTES

1. Ian Howell, "Misreading the Science: Vocal Treatises, Vowels, and a New Framework for Understanding the Female *passaggio*," unpublished term paper MHST902, New England Conservatory of Music, 2014.
2. Cornelia Fales, "The Paradox of Timbre," in *Ethnomusicology* 46, no. 1 (2002): 58.
3. Reinier Plomp, "Experiments on Tone Perception" (PhD diss., Institute for Perception RVO-TNO, 1966), 131–32.
4. William Hartmann, "Pitch, periodicity, and auditory organization," in *Journal of the Acoustical Society of America* 100, no. 6 (December 1996): 3494–95.
5. Ian Howell, "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color" (DMA diss., New England Conservatory of Music, 2016), 29.
6. Plomp, "Experiments," 132.
7. Robert Cogan, *Music Seen, Music Heard: a Picture Book of Musical Design* (Cambridge: Publication Contact International, 1998), 110.
8. Howell, "Parsing," 29.
9. Howell, "Parsing," 39.
10. Howell, "Parsing," 73; Ian Howell, "Necessary Roughness in the Voice Pedagogy Classroom: The Special Psychoacoustics of the Singing Voice," *VOICEPrints* (May/June 2017): 5.
11. Howell, "Parsing," 31
12. David J. M. Weenink, "Formant Analysis of Dutch Vowels from 10 Children," in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 9 (1985), 45–52.
13. For clarity, this is "close" [kloɔ̃s] that is the opposite of *far*. It refers to the typical distance of the tongue dorsum from the roof of the mouth.

10

How Pitch, Auditory Roughness, and Tone Color Intersect

Theory and Application

Chapters 8 and 9 explored the ideas of brightness and the tone color properties of pure tones. I believe that coming to understand the timbral qualities of these *perceptually uncomplicated* sounds is a necessary first step toward grasping the differences between the co-presenting, qualitatively opposable spectral features of a complex sound. Additionally, chapters 5 through 7 were sequentially arranged to build your understanding of pitch resolution and auditory roughness. These three potential timbral qualities of a sound—pitch, auditory roughness, and tone color—are academically interesting in isolation. However, I do not think that any one of them sufficiently captures the timbre of a voice in an actionable manner. I argue that the utility of these percepts becomes clearer once you understand how they intersect as a part of the complex interplay of timbral qualities in a voice. This chapter will explore many examples from real singers. It also offers a rule-of-thumb framework to help you read perceptual characteristics into a spectrum or spectrogram.

What follows may be summarized in the following way:

In much of the singing range we can *hear* formants. They are not just measurable spectral features. Those spectral features point to real, physical pressure changes that stimulate our ears.

Those formants may be “moved” from dark to bright in tone color (low to high frequency), especially the first and second formant (F_1 and F_2).

The faster the formant is—relative to the fundamental frequency of vocal fold oscillation—the buzzier and noisier it is likely to be.

Even if your aim as a voice teacher is to use only your ears—and that is my aim; I use no computer when I teach—spending some time with the pass-filters in a program like VoceVista Video Pro or Praat can powerfully train and condition your ears.

THE TIMBRE OF A FORMANT AS PITCH CHANGES: A THOUGHT EXPERIMENT

Join me in a thought experiment. I suggest that a pure tone at F7 (~2.8 kHz) has an |~i| tone color. Listen to it in isolation and you will likely notice its brightness. Play it into your mouth with a small speaker (for example, from an iPhone) and you will notice it “pops” into resonance when you shape your vocal tract for [i]. However, by definition, that pure tone has neither auditory roughness nor a pitch-less, unresolved portion of its spectrum. When strong energy around 2.8 kHz exists in a spectrum, we must consider the intersection of that tone color with the auditory roughness and pitch resolution of the formant it is part of. Figure 10.1 illustrates this. All four images represent sounds with very similar tone colors in the peak centered on F7 (~2.8 kHz) but with differing degrees of auditory roughness and pitch resolution. You can clearly visualize that spectral component in all four waveforms.

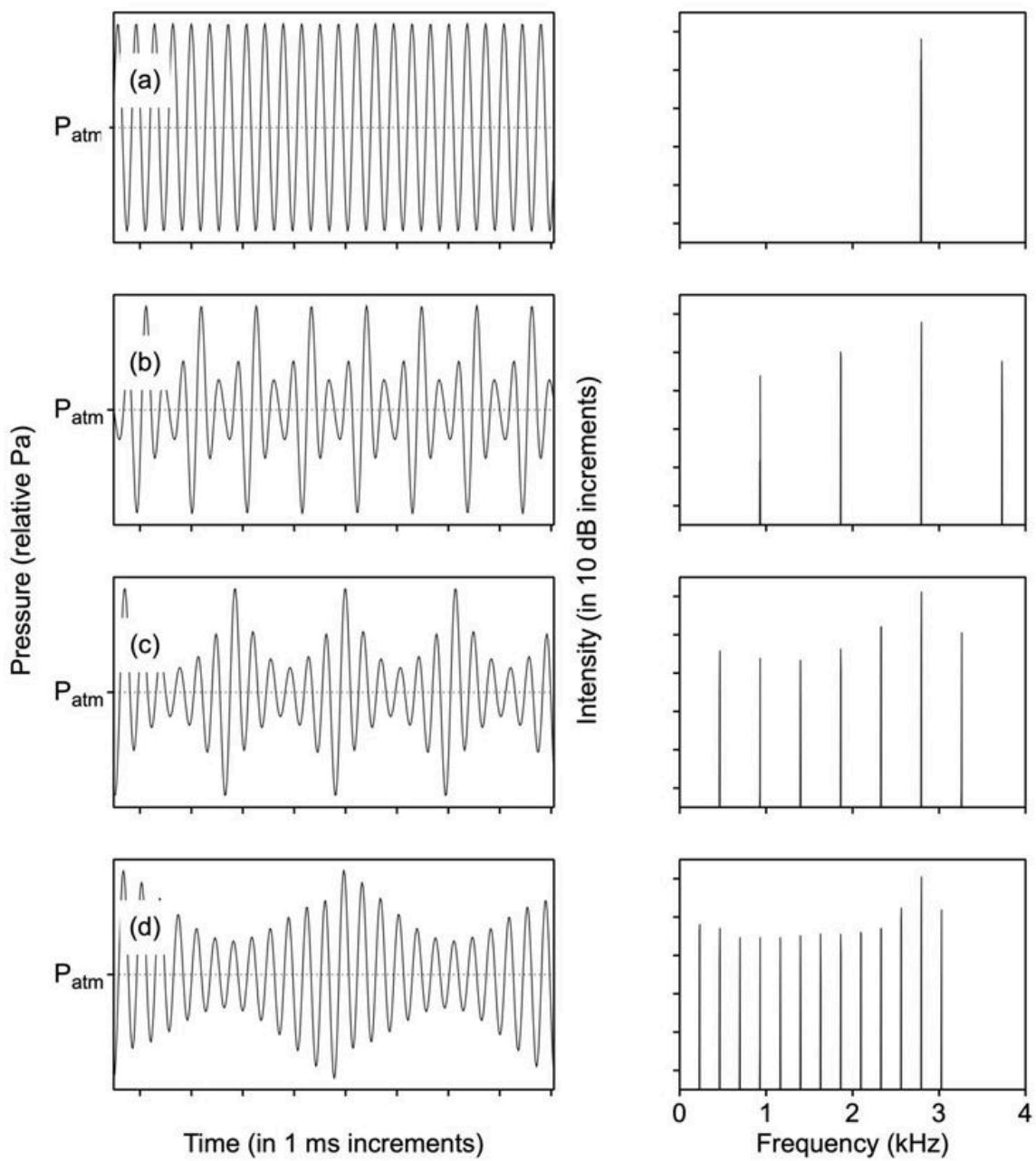


Figure 10.1 Four images (a–d) showing waveforms (left) and spectra (right) that all have a strong peak centered on ~ 2.8 kHz. (a) shows a pure tone at 2.8 kHz. (b) shows a peak at ~ 2.8 kHz on the third harmonic ($3f_0$). (c) shows a peak at ~ 2.8 kHz on the sixth harmonic ($6f_0$). (d) shows a peak at ~ 2.8 kHz on the twelfth harmonic ($12f_0$) which is unresolved and very rough (buzzy). Source: Synthesized by author in Madde.

Figure 10.1 (a) shows a sine tone where the frequency has an |~i| tone color. We already know that this sample would elicit no auditory roughness because it is $< 5f_0$, and there is no noisy, pitch-lessness because it is $1f_0$. It is also below the 5 kHz threshold, but all these samples are. Figure 10.1 (b) shows a peak at ~ 2.8 kHz centered on what would be the third harmonic ($3f_0$) of a periodic tone with a fundamental equivalent to ~ 933 Hz (around B♭5). In this sample, the strong |~i| quality elicits no auditory roughness because it is $< 5f_0$ and resolves into the pitch because it is $< 9f_0$. Figure 10.1 (c) shows a peak at ~ 2.8 kHz centered on what would be the sixth harmonic ($6f_0$) of a periodic tone with a fundamental equivalent to ~ 466 Hz (around B♭4). In this sample the strong |~i| quality is resolved into the pitch because it is $< 9f_0$ and rough (buzzy) because it is $> 5f_0$. Figure 10.1 (d) shows a peak at ~ 2.8 kHz centered on what would be the twelfth harmonic ($12f_0$) of a periodic tone with a fundamental equivalent to ~ 233 Hz (around B♭3). In this sample the strong |~i| quality is unresolved from the pitch (noisy) because it is $> 8f_0$ and very rough (buzzy) because it is well above $5f_0$.

Note that in samples (b) through (d), I included equal intensity harmonic energy on either side of the target frequency of 2.8 kHz. I also filled in the lower harmonics of whatever harmonic series fit. For example, this is what I meant by suggesting that in (b), ~ 2.8 kHz would be $3f_0$ of a fundamental at B♭5. It is worth thinking through why I made these choices. Remember that harmonics are a way to characterize faster moving aspects of a periodically repeating pattern. This pattern is characterized by the faster pressure patterns of the vocal tract resonances that repeat within the slower pressure change of the fundamental. Without those additional harmonics in the synthesized sound, there would be no periodic repetition of that slower pattern, and these formant peaks would elicit no qualities of roughness or pitch resolution. In all three cases, there is a fast ripple at ~ 2.8 kHz that rides along a slower pulse at the frequency equivalent to the pitch percept. If you would like to explore this or generate your own custom files, see Lab #20.

Those samples clearly sound synthesized. So what does this sound like in a real human? Fortunately, many singing voices have a spectral peak in the range of ~ 2.8 kHz (around an $|\sim i|$ tone color). It does not matter whether you have a sample that precisely centers on that frequency or not, as $|\sim i|$ covers a wide frequency range. By recording a range of pitches and pass-filtering a region surrounding the center frequency of the nearest peak, you will hear that the tone color of the sample remains similar even as the pitch resolution and auditory roughness changes. This will be true regardless of the vowel you choose, although note that vowels like [u] and [o] tend not to concentrate much energy in that higher $|\sim i|$ range. Sing a glide from high to low and you will hear auditory roughness and pitch-lessness emerge in the pass-filtered frequency band as $1f_0$ falls, while the tone color remains similar (see [figure 10.2](#) for an example of this). See Lab #21.

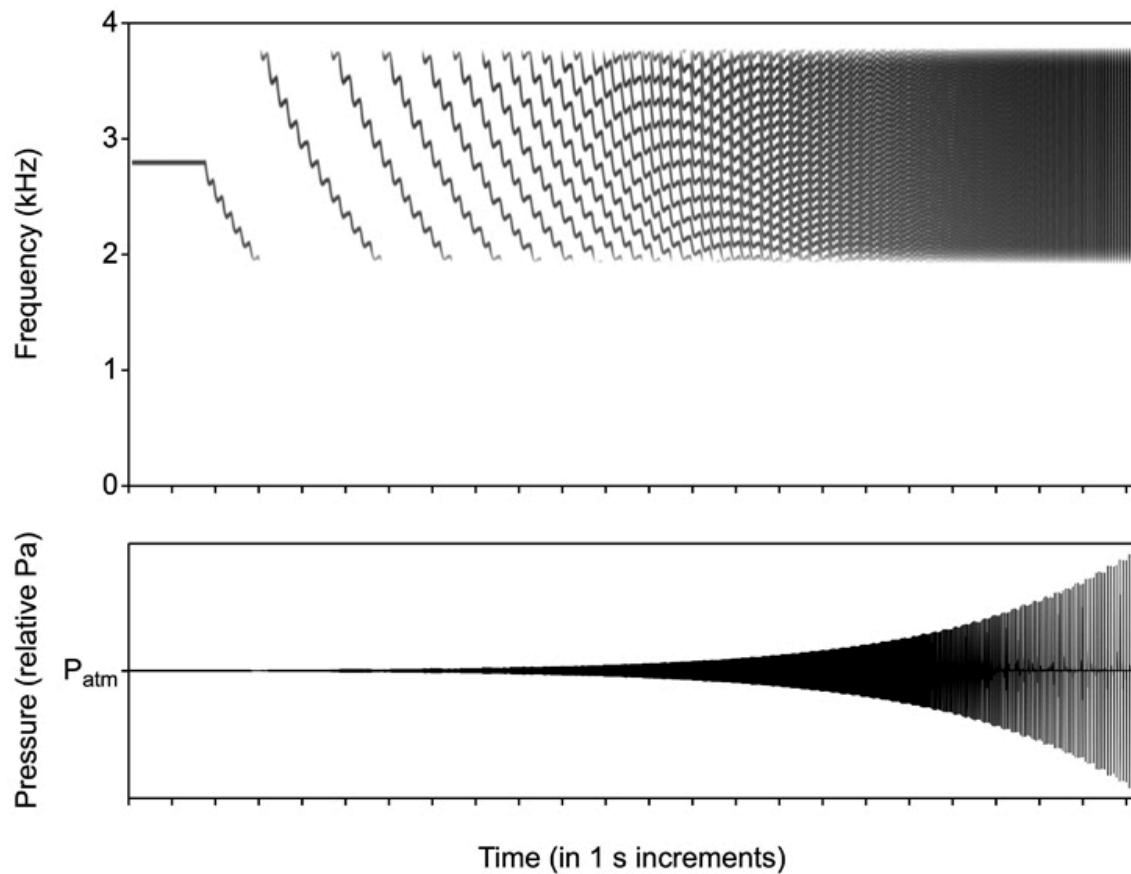


Figure 10.2 (Top) spectrogram and (bottom) waveform of a scale from $1f_0 = \sim 2.8$ kHz to ~ 12 Hz, pass filtered between 2 kHz and 3.7 kHz with a 100 Hz smoothing. Source: Synthesized by the author in Madde.

Additionally, the timbre of a complex tone tends to be temporally variable; it varies over time. The changes undergone by a pure tone are limited to amplitude or frequency. A complex tone may independently vary the presence of energy at multiple frequencies in the spectrum, and we know that the ear responds differently to these different portions of the spectrum. To return to figure 7.9 from [chapter 7](#), a sung *decrescendo* tends to remove higher frequency energy from the spectrum quicker than it removes lower frequency energy. Those high-frequency and low-frequency regions of the spectrum may elicit different amounts of auditory roughness and pitch resolution, but they also trigger contrasting tone colors. They literally sound different from one another. The percepts of such sounds are formed not just at the intersections of these multiple timbral qualities but also in the way or ways they change over time.

THE INTERSECTION OF AUDITORY ROUGHNESS, PITCH RESOLUTION, AND TONE COLOR IN A SINGLE FORMANT

As explored above, the application of ASTC to complex tones is . . . complex. It must be acknowledged that much of the pressure pattern of a sung sound is not filtered by the ear into the harmonics displayed in a spectrum or spectrogram. Even if harmonics were “real” and persistent physical phenomena, the critical bands of human hearing suggest that most harmonics could not be heard separately anyway. This means one may not presume that the individual harmonics seen on a spectrogram exist as separate percepts in the way that individually generated pure tones do. Some harmonics absolutely represent persistent oscillations that can be separately perceived. But they are the exceptions. In practice, we should look at the formant structures—not the harmonics—in the spectrum to truly grasp what we perceive. Those formant structures will also exhibit additional timbral qualities beyond tone color arising from the nature of their complexity.

So how can we talk productively about how tone color intersects the additional timbral qualities of pitch resolution and auditory roughness? Accepting that there is always greater nuance, in simpler terms, pitch resolution and auditory roughness move with the fundamental frequency of vocal fold oscillation. We can imagine tone color as the space they move through. This means that different formants in the spectrum—each with its own different and varying degrees of pitch resolution and auditory roughness—correspondingly move through different tone colors as pitch changes. The formants represented by rising and falling harmonics in a spectrum or spectrogram potentially fill in a wide range of possible tone colors. Remember, I am positioning tone color as an effective way to understand the qualitative properties of the frequency domain, independent of a given position within a harmonic series. For example, it would be nonsensical to suggest that $5f_0$ in general has a single, specific,

identifiable tone color. It does not. However, $5f_0$ at a specific frequency does.

Next, consider the perceptual qualities of auditory roughness. We have learned that this percept is usefully understood through its relationship to the fundamental of a periodic sound. That is to say, for most of the singable pitch range, roughness begins above the fifth harmonic ($5f_0$). In general terms, as the fundamental rises or falls in a pitch space (typically by increasing or decreasing in frequency), roughness tends to track similarly. A more nuanced view would capture that as pitch rises above the treble staff, roughness begins higher in the harmonic series, and as pitch falls below the bass staff, roughness begins lower in the harmonic series.¹

With the understanding that the de facto ceiling of 5 kHz (> D \sharp 8) acts as a sort of aural threshold, pitch similarly merits additional nuance. As a brief aside, the piano's highest note is C8, at 4186 Hz; the notes of a piano stop just below that upper pitch percept threshold. The highest notes on a violin (A7) and concert harp (G7 or G \sharp 7) are similarly just below that same threshold. This is no coincidence. Why waste building materials generating playable sounds no one would hear as a pitch? Most periodic energy above this 5 kHz threshold tends to not resolve into the pitch, regardless of its location in the harmonic series. But again, for *most* of the range singable by humans, ordinal position within the harmonic series points to pitch resolution. Again, the simple rule of thumb is that harmonics one ($1f_0$) through eight ($8f_0$) most significantly contribute to the pitch percept, while higher harmonics contribute an increasingly noisy quality.² You may alternately think in terms of the slower, less rapidly damped resonances of the vocal tract contributing to the pitch percept, while the quickly attenuated faster resonances sound like buzzy noise instead.

Now consider this thought experiment. A singer sings a sound with a strong formant that appears on a spectrogram as the fifth through ninth harmonics ($5f_0$ – $9f_0$). We will assume that this formant is so prominent that it dominates the timbre (see [figure 10.3](#)). Crucially, we will also

assume that as $1f_0$ rises and falls, that the peak similarly rises and falls in frequency; it remains perfectly associated with $5f_0$ – $9f_0$. In the time domain, this would suggest that resonance oscillates increasingly faster as the pitch period shortens.

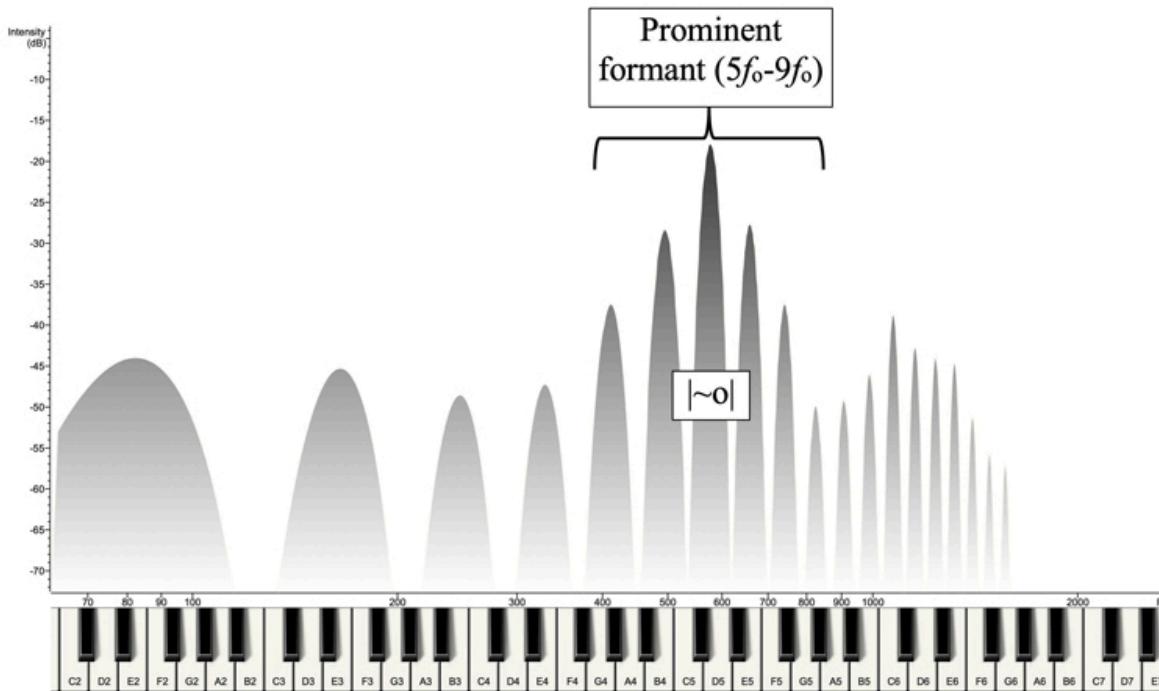


Figure 10.3 Spectrum of a sound with a prominent formant centered on the seventh harmonic ($7f_0$). Source: Synthesized by author in Madde.

Based on the rules of pitch and roughness as explained in [chapter 7](#), we would expect this part of the spectrum to be buzzy ($> 5f_0$) and resolved into the pitch ($< 9f_0$). As pitch rises and falls, these percepts follow. However, because this formant is moving higher and lower in frequency as the fundamental moves, the *tone color* of this portion of the spectrum changes as pitch rises and falls.

Consider this portion of the spectrum ($5f_0$ – $9f_0$) in isolation for a moment. When the fundamental is D2, this formant sounds somewhat like an American English [o] (corresponding to the |~o| ASTC of a pure tone at the frequency of D5, or ~587 Hz), because the prominent center of this formant is approximately centered on that frequency. In practical terms,

we do not need to measure the distribution of the frequency energy in this formant to estimate the ASTC. The strong harmonic in the middle is easily seen in the spectrum. Raise the fundamental of this harmonic series while preserving the intensity of each harmonic (the formant also rises), and the color of this peak will perceptually *open* as though moving through incomplete versions of [ɔ], [ɑ], [a], [æ], [e], and finally [i]. If you do this with an awareness of the ASTC labels, you will notice that the tone color changes along that continuum. However one labels this phenomenon, it is easy to demonstrate these contrasts along the tone color continuum. The auditory roughness and pitch resolution of this rising formant remain reasonably similar for most of the potential pitch range, eventually turning into buzzy, pitch-less noise. Most voice teachers I have demonstrated this to liken that final sound to the sound of crickets. See Lab #22.

Next, imagine that instead of rising with the fundamental, the formant in figure 10.3 instead persists at the same frequency while the pitch rises. In the time domain, this would suggest that the resonance in the vocal tract continues to oscillate at the same frequency regardless of how fast the vocal folds oscillate. Since the frequency of that resonance remains constant, and thus the frequency center of the resulting formant correspondingly remains constant, the tone color does not change much. Instead, when the pitch is low enough, that tone color will sound like buzzy noise. As pitch rises, that tone color will resolve into the pitch. A little higher and that tone color will lose its buzziness. Once the fundamental aligns with the resonance, it will sound substantially like a pure tone. This is the reverse of the process explored in figure 10.2. See Lab #23.

Strange things start to happen to the tone color of this formant as $1f_o$ begins to rise close in frequency to the frequency of that resonance. Remember, as a physical oscillation of the air in the vocal tract, the frequency of that resonance points to how quickly it completes one cycle oscillating from low to high and back to low pressure. At the end of each oscillation, that low-pressure front returns to the air mass just above the glottis. When the pitch is low (period is long) relative to the frequency of

the resonance, for example the arrangement shown as a spectrum in figure 10.3, that resonance has already traveled forward and backward in the vocal tract many times. As such, it would be subject to more significant damping, and there may not be much interaction of that resonance and the next supraglottal impulse. If instead that resonance pattern only unfolds three times, that resonance will undergo far less damping.

To return to Titze's analogy of a child on a swing, if you only push them once every seven swings, their oscillation will have dampened (diminished in amplitude) to the point that your push may apply more force to the swing than their swinging does to your hands. If instead that oscillation only unfolds three times, it will be far less impacted by damping and the average amplitude remains higher. Repeatedly interrupting that oscillation two and a half times through its pattern—when the kid has yet to swing back to you, or the returning low-pressure front has yet to make it back to the vocal folds—may noticeably impact the average amplitude of that ongoing oscillation and, by analogy, the amplitude of the radiated sound. If you try to push them halfway through their first swing, you may get hurt. Fortunately, the analogy breaks down for singing here, but the average amplitude of a resonance that oscillates at $1.5f_0$ is similarly weaker.

Aligning the supraglottal impulse with the return of that low-pressure front allows the oscillation of the resonance to transfer its motion across glottal cycles, which helps to establish a standing wave in the vocal tract. This allows acoustic energy to accumulate in the vocal tract and helps create the percept that strong aspects of the continuous sound emerge from a discrete, periodic process. In singing, as in pushing a child on a swing, amplitude matters. A mistimed push at the two-and-a-half-swings mark makes a bigger difference the harder you pushed the child the first time. In this way, the acoustical relationships of consequence tend to go away when singing quietly, and these beneficial relationships become consequential when singing loud, high pitches.

Back to our thought experiment. As pitch rises closer to the stable formant frequency, consequential mistiming occurs between the end of

the resonance and the next supraglottal impulse. In a spectrum we see these mistimed relationships as harmonics straddling the imagined resonance that is between them. The effect is a weaker, more neutralized sound. In the time domain, the resonance starts every period oscillating at the same frequency that it would if there were a perfect alignment with a later supraglottal impulse. It is just *rudely* interrupted by the vocal folds at an inopportune time. The spectrum displays this as *weaker harmonics to either side* because the Fourier transform preferentially represents repeated patterns within the pitch period as *harmonics of that fundamental*.

Thus, as pitch rises and that resonance remains the same, you will start to hear the timbre alternate between a clear vowel sound and a more muted or neutral version of that vowel sound. This typically becomes noticeable when the resonance is between $3f_o$ and $4f_o$, and even more so when between $2f_o$ and $3f_o$. The timbral transition from f_{R1} aligning with $2f_o$ and that resonance sitting between $1f_o$ and $2f_o$ —meaning it is interrupted by a pressure drop as it is partway through its second pressure cycle—is so strong that Bozeman calls this transition the “pitch of turning.”³ He further labels the mistimed relationship $1f_o < f_{R1} < 2f_o$ after the historical term *voce chiusa* (closed voice or close [kloɔ̄s] timbre), and the sound prior to that mistiming ($2f_o < f_{R1}$) after the similarly historical term *voce aperta* (open voice or open timbre).⁴

THE TONE COLOR OF “FLAT” VERSUS “POINTY” FORMANTS

One can predict what the tone color of a spectral peak will be if an integer multiple of its resonance strongly aligns with the frequency of vocal fold oscillation. For example, every one, two, three, etc., times through the oscillation of the resonance aligns with the next glottal contacting event. In a spectrum or spectrogram, this would appear as a formant peak with a high-intensity harmonic at its center. However, when confronted with a resonance that misaligns with the frequency of vocal fold oscillation, that formant flattens out. No single harmonic appears as a clear central frequency, so we presume that the resonance is *hidden* in the spectrum between two harmonics.

To explore the tone color of such a formant, you can use a rule of thumb that leverages a measurement called the *spectral centroid*. In Praat it is named the *centre of gravity*, and it is assigned a function in the Query menu of a Spectrum object. (Note the British English spelling of *centre*.) This measurement returns an amplitude-weighted average frequency (Hz) for a spectrum or portion of a spectrum. Essentially, if you have a formant with a strong central harmonic, the frequency of that harmonic *is* the spectral centroid for that part of the spectrum. If you have a formant consisting of two harmonics of equal intensity, you will most likely hear the tone color of what a pure tone between them would be.

Here are some examples of this: Combine a pure tone at 440 Hz (A4) and another at 880 Hz (A5). To calculate the spectral centroid, you add those frequencies and divide by the number of pure tones. Were those pure tones different in amplitude, the calculation would be more complex. Here, $440 + 880 = 1,320$; $1,320 \div 2 = 660$. Thus, the unified tone color of those two equal-amplitude pure tones sounding simultaneously will be |~o| (see again [figure 9.1](#)). Repeat this with pure tones at 523 Hz (C5) and 784 Hz (G5). The spectral centroid is once again ~660 Hz, which still sounds like |~o|. Even if the first tone is ~623 Hz and the second ~693 Hz, the spectral centroid is again ~660 Hz and the tone color will be |~o|.

This last sample will be so rough that it will remind you of an emergency alert tone. It definitely gets your attention. These three pairs of pure tones represent harmonics of equal intensity at $1f_o$ and $2f_o$ of A4, $2f_o$ and $3f_o$ of C4, and $9f_o$ and $10f_o$ of C#2, respectively. See Lab #24.

It is also possible to calculate what the spectral centroid of pure tones of *differing* amplitudes would be. If, for example, you saw a formant in a spectrum that featured a strong lower harmonic and a weaker higher harmonic, the intensity of the lower harmonic will *pull* the resulting average frequency lower. This triggers a correspondingly darker tone color, which aligns better with the ASTC of that lower frequency yet higher intensity harmonic. However, we now run into a common challenge related to reading spectra and spectrograms. The *equal loudness contours* introduced in [chapter 3](#) suggest that the human ear will preferentially overrepresent the intensity of certain frequency ranges that fall between around 1 kHz and 3.5 kHz, and that the degree to which this happens depends on the overall intensity of the pressure pattern at the ear. So, no matter what, the objective measurements of the relative intensities of the harmonics shown on a spectrum or spectrogram will be further reshaped by the process of auditory transduction. The complexity of this reshaping in both the frequency and intensity domains makes it virtually impossible to precisely measure our percept of the spectral centroid. Fortunately, no practical experience of the tone color of a formant requires careful measurement or calibration. It is enough to be *close* in a voice lesson.

RULES OF THUMB

To summarize, pitch and roughness generally relate to the *position of a formant within the harmonic series*. Tone color, however, is related to the central *frequency* of that formant. The harmonics we see in a spectrum or spectrogram move freely through that tone color space. This points to the following three complementary patterns at the intersection of pitch, auditory roughness, and tone color:

1. If the fundamental frequency changes while a resonance stays the same, the roughness and pitch resolution of the resulting formant may change. Once the resonance sits between low-integer multiples of the frequency of vocal fold oscillation, the vowel quality may neutralize.
2. If the fundamental frequency stays the same while the resonance speeds up or slows down (raises or lowers its frequency), the resulting formant may move in and out of roughness and pitch resolution as it changes tone color.
3. If the resonance speeds up its frequency proportional to an increase in the frequency of the fundamental (sometimes called *formant tracking*), the roughness and pitch resolution of that formant will remain basically stable as its tone color changes.

Some Examples of These Intersections

The above framework helps to explain easily observed phenomena. For example, as pitch rises, the sound of a voice qualitatively simplifies. A soprano's high F6 stimulates an objectively less-complex percept when compared to a bass's F2. The F6 is likely devoid of any auditory roughness

or unresolved harmonics, and its fundamental oscillates faster than every tone color below |~a|.

Midrange examples also benefit from a similar thought experiment. Consider the four examples in figure 10.4 (a–d) showing waveforms (left) and spectra (right) that were generated by the Madde voice synthesizer. All four have the same set of resonances, which correspond to an average [a]. At G2 (a), there are separate formants (F_1 and F_2) related to the first and second resonances (f_{R1} and f_{R2}). Those resonances also fall above the fifth harmonic ($5f_0$), so the vowel's defining quality is buzzy. Much of this spectrum is unresolved and contributes noise rather than pitch. An octave higher at G3 (b), there are still two vowel formants, although they are less buzzy and less noisy. An octave higher again at G4 (c) is markedly different. Gone are the separate formant peaks for each of the lowest two resonances. Because the pitch is higher, that portion of the spectrum is now entirely pure, with auditory roughness reserved for higher formants in the spectrum. All the harmonics shown resolve into the pitch. An additional octave higher at G5 (d) shows scant evidence that the complex peaks present at G2 were ever present. Gone too is any tone color below the primary peak of $1f_0$. This demonstrates that voices obligatorily “simplify” in timbre in very specific ways as pitch rises, and that at least for this vowel, much of that change occurs by the bottom of the treble staff.

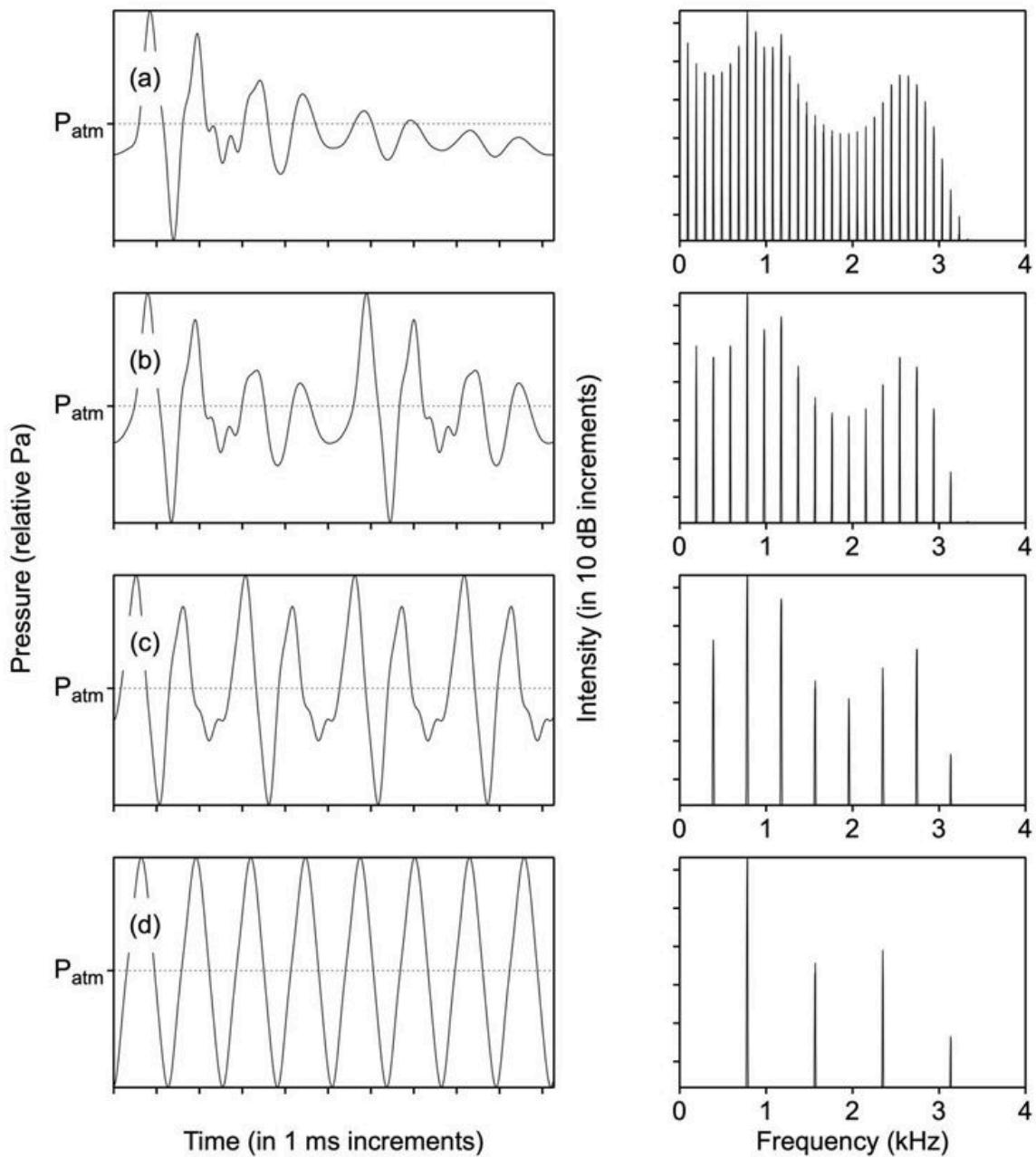


Figure 10.4 Waveforms (left) and spectra (right) of the [a] vowel at four pitches, G2 (a), G3 (b), G4 (c), and G5 (d). Timescale is preserved in waveforms. Source: Synthesized in Madde by author.

Note too that the waveforms (figure 10.4, left) all have the same timescale. Moving from G2 (a) to G3 (b) halves the period duration. Moving from G3 (b) to G4 (c) and G4 (c) to G5 (d) similarly halves the

period duration. Since the resonances have not changed, the waveform always starts out the same. The pitch determines how far through that pattern the resonance continues to oscillate.

THE OBVIOUS TRUE FUNDAMENTAL

I sang professionally as a countertenor, a voice type that spends much time on the treble staff. Thus, my initial exploration of the psychoacoustics of the singing voice focused on registration on and above the treble staff. Consequential timbral changes occur in this pitch range, many of which characterize voice type and genre. At the time (the first two decades of the twenty-first century), much of the literature surrounding voice acoustics centered on cisgender male voices singing as tenors and baritones (and sometimes basses) in classical genres. This work was disproportionately preoccupied with the acoustic output of tenors, especially tenors singing “money” notes. Thankfully this is changing, but work remains to be done to understand treble voices with equivalent thoroughness.

Based on our explorations so far, consider the following: Around approximately C4 (~262 Hz), the ASTC of the fundamental ($1f_0$) is $|\sim u|$, while the second harmonic ($2f_0$) crosses a perceptual threshold from $|\sim u|$ to $|\sim o|$. Below C4, both $1f_0$ and $2f_0$ fall entirely within $|\sim u|$. This means the fundamental ($1f_0$) potentially conveys a unique tone color starting at and above C4. Whether it is expressed or not depends entirely on the rest of the spectrum.

I repeat that sine waves are an interesting idea, but they rarely exist in nature. But in a way of thinking, the color of $1f_0$ at and above the pitch C4 is an exception, much like a clear harmonic in overtone singing. A strong $1f_0$ at and above C4 is potentially perceptually separable from the rest of the spectrum, especially if the pitch and vowel give it high amplitude as a resonance. See again Figure 8.2 for an example of a persistent, high-intensity $1f_0$ even as vowels change. The fundamental in common between that [u] and [i] is perceptually separable, will never exhibit auditory roughness or unresolved noise, and perceptually behaves like a pure tone. That is to say, we may learn to hear its simple ASTC. No other harmonic tends to generate this percept so dependably, although I will share counterexamples later in this chapter.

I call this *the obvious true fundamental*. *Obvious* refers to how its tone color contrasts with the rest of the spectrum. *True fundamental* refers to the fact that while it only contributes a portion of the pitch percept, this contribution functions as the tone color of a pure tone equivalent to the pitch. The remainder of the pitch percept is then composed of missing fundamentals, with brighter tone colors generated by faster formants. Because classical trebles often shape their vocal tract to continuously align their lowest vocal tract resonance with the first harmonic as pitch ascends—remember, this resonance strategy is called *whoop*⁵ or *hoot*⁶—this harmonic is often significantly boosted to become more intense than the rest of the spectrum.⁷ This means that regardless of the additional, within-period oscillations of resonances, the pitch period is characterized by one large ripple from low to high pressure, and then back to low.

Richard Miller (2004) similarly suggests, "Among prominent female artists, when they are singing in [the] upper range, the first formant and the fundamental are often enhanced and exhibit increased acoustic energy in the lower portion of the spectrum."⁸ This causes a portion of the total vowel percept to be significantly defined by the simple ASTC of the *obvious true fundamental*, rather than by the perceptual qualities of more-complex formants that may be found in the lower pitch range of the voice.

As pitch rises through and above the treble staff, we recall that the absolute spectral tone color of a pure tone would pass through |~u|, |~o|, |~ɔ|, |~ɑ|, and (at the upper extreme) |~a|. Endogenous testosterone puberty singers may spend their entire life singing with their fundamental in the |~u| tone color. This means that treble voices singing with a strong fundamental will experience timbral shifts around C5 (shift from |~u| to |~o|), F#5 (|~o| to |~ɔ|), C6 (|~ɔ| to |~ɑ|), and to |~a| by F#6, regardless of their target vowel. These obligate tone color changes do not necessarily characterize the complete timbre of that treble singer, but they will absolutely color those transitional pitch regions. Richard Miller (2000) places a soprano's acoustic registration transitions at approximately C#5, F#5, and C#6, which maps well to this schema.⁹

Miller gives mezzo-sopranos and contraltos slightly lower transitional pitch points, perhaps implying that some aspect of this registration is physiological. For an equal amount of physical ease, the pitch range of the first vocal tract resonance may lie lower for lower voice types, making corresponding resonance tuning transitions easier at slightly lower pitches. However, the similarity of these pitches to the obligatory shifts in the ASTC of the first harmonic (independent of vowel) may suggest that these particular registration points rest on a significant psychoacoustic component as well. The expectation of how a vowel percept will change as pitch rises may then be a combination of perceptual and physical considerations.¹⁰ You may wish to record treble singers and pass-filter the fundamental to note the way in which the resulting tone color will change from the bottom to above the top of the treble staff. See Lab #25.

Taken with predictable changes in auditory roughness and pitch resolution, the change in the tone color of the fundamental ($1f_o$) appears to explain paralinguistic changes in the *timbre* of a treble singer as pitch rises. Looking at registration in this way accommodates the reality that lower-timbred (dark) voices do not sound the same as higher-timbred (bright) voices as they execute comparable acoustic registration events. For example, a contralto singing an [o] at an A4 may enter *whoop* registration ($1f_o : f_{R1}$), but a soprano may delay this shift until D5. Those obvious true fundamentals will elicit different tone color percepts at those two pitches; correspondingly, these voices will sound different. The contralto's fundamental will contribute a strong |~u| at A4 that in a soprano voice will be much weaker at the same pitch and vowel combination. A soprano will contribute an |~o| when they reach *whoop* registration ($1f_o : f_{R1}$) around the pitch D4. This maps intuitively to the idea that a darker-timbred voice achieves that darkness from profiling the tone colors of their relatively lower frequency formants.

The lens of ASTC is tied to frequency. From this perspective, higher sung pitches are perceptually different from lower sung pitches in a way that deserves consideration. Practitioners working primarily with lower voices may not be aware of this. Practitioners who conceive of the soprano voice as functionally identical to a tenor voice an octave higher

may not be primed to notice this. It is my hope that such practitioners join other subsets of voice educators in engaging these phenomena using the ASTC framework. This material has direct, wide-ranging, demonstrable application in the singing voice studio.

A COMBINED PERCEPTUAL MODEL

The addition of ASTC values completes the model introduced in [chapter 9](#). In figure 10.5 (top), note that the intersection of ASTC, pitch resolution, and auditory roughness creates a complex percept for the G3 from figure 10.4 (b). The vowel formants are discrete, the singer's formant is buzzy and unresolved, and two harmonics populate the |~u| tone color range. The intersection of ASTC, pitch resolution, and auditory roughness is different for the G4 in [figure 10.5](#) (bottom). The two vowel formants now produce a single peak with a darker average tone color than the second formant (F_2) of the G3. The singer's formant now falls within the resolved portion of the spectrum, and the fundamental is the only harmonic with the |~u| tone color, which affords it an authentic ASTC label. If you pass-filter each annotated spectral region in both samples, you will hear the qualities I have outlined here.

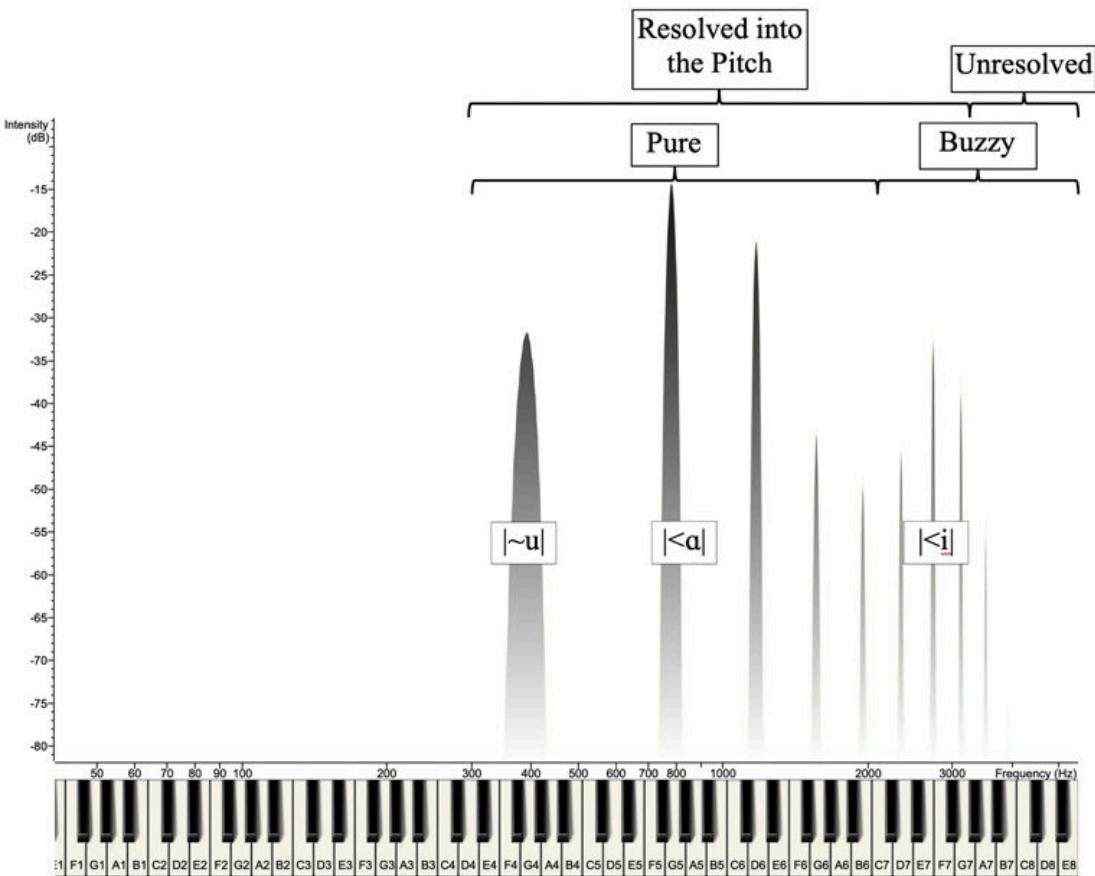
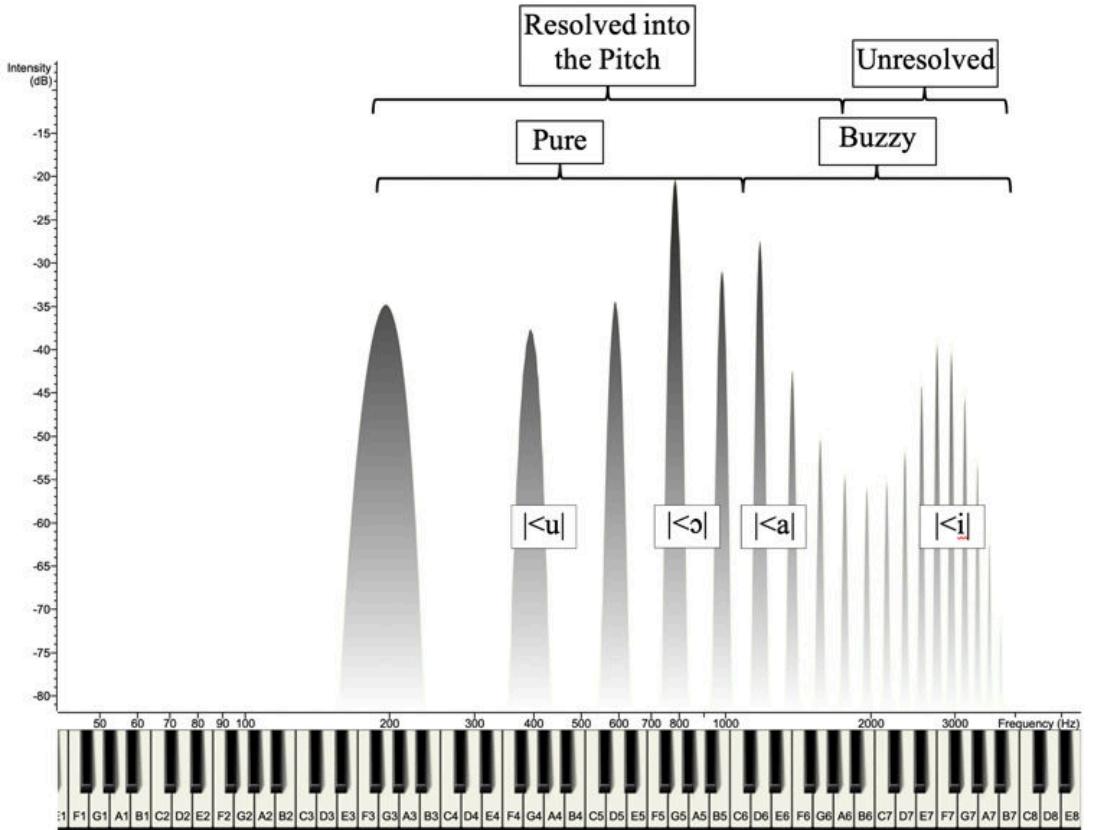


Figure 10.5 A perceptual annotation of [a] at G3 (top) and G4 (bottom). Source: Created by author.

Note that the color of a cluster of harmonics in this image is denoted with a “<” rather than “~.” This is intentional, because **ASTC values only hold for pure tones**. Once a cluster of harmonics is averaged together into a complex formant, the tone color is something more complex yet still *less than* a complete vowel. Engage the analysis at this level or granularity or not. As a practitioner, it is enough to simply encounter the tone color of the formant in question.

In the next section I will explore several approaches to singing featured in classical and contemporary genres. Please try to apply the model I propose, and visually annotate the spectra with pitched/pitch-less and pure/rough divisions. Consider the tone colors and position within the harmonic series of the various formants as they appear. Are they more like figure 10.4 (a) and (b), with two identifiable vowel formants (F_1 and F_2) and many unresolved harmonics? Or are they more like figure 10.4 (c) and (d), with fewer unresolved harmonics and a simplification of the vowel formants?

PERCEPTUAL SIGNIFIERS OF SOME COMMON RESONANCE STRATEGIES

Having spent time exploring the various rules of thumb for visually annotating a spectrum or spectrogram, and having considered the timbral difference between a formant with a strong central harmonic (a well-timed resonance) versus a formant without that strong central harmonic (a poorly timed resonance), I would like to explore several real-world recordings where singers chose to time their resonances well. These are examples that invite us to extend ASTC labels to inform the tone colors of the resulting formants. This typically only happens when there is a strong resonance, with minimal damping, oscillating in the vocal tract for the entire pitch period. This exists in at least three ways of singing:

1. Overtone singing. When the formant's bandwidth is so narrow (the damping of the resonance is so slight) that the harmonics surrounding the formant peak are significantly weaker, we hear that faster ripple as a separate pitch percept. In this edge case, we do not typically hear auditory roughness or pitch-lessness below 5 kHz from that formant, regardless of its place in the harmonic series.
2. Another way this occurs is with strong energy at $1f_0$ on and above the treble staff. This is a common strategy to build power in a classical treble voice.
3. A third way features formants that fall somewhere between overtone singing and the kind of spectral peaks typically found in speech—in other words, well-populated with harmonics in a spectrum. This approach is to sing with a strong, faster resonance

that still aligns with the period of $1f_0$ and exhibits little damping within that period, whether or not that formant rises to the percept of an overtone. This resonance strategy can be found across genres. Singers typically use this strategy on high-energy, higher-pitched “money” notes. In a spectrum, we typically see a strong $2f_0$ or $3f_0$ (or possibly $4f_0$ or $5f_0$), and the percept of the vowel simplifies in a characteristic way.

The above approaches to navigating higher pitches all feature strong acoustic energy oscillating within the vocal tract. These oscillations have timbral qualities determined at least in part by the tone colors of those strong standing waves. To refresh your memory regarding overtone singing, you can refer to [chapter 6](#). In the following two sections I will discuss the second and third approaches shown above.

The Tone Color of a Strong $1f_0$: Solutions for the Treble Staff and Higher

Classical treble singers and many contemporary singers with a warm sound commonly generate a strong first formant that coincides with $1f_0$ on and above the treble staff. This favors the warmer, pure regions of the spectrum, which both contributes to the timbre and creates a resonance potentially strong enough to act as a driving force for phonation. Or at least a potential driving force that assists the generation of the supraglottal impulse. The degree of such nonlinearity is related to the amplitude of oscillation of that resonance. I will explore examples from two classical singers here, but go listen to the first 1:08 of Whitney Houston’s recording of “I Will Always Love You.”¹¹ She features a persistent $1f_0$ throughout, which contributes a pure |~u| quality

regardless of the vowel. More on the hazards of studying commercial recordings in the next section.

Vowels with a lower first resonance (f_{R1}) such as [u] and [i] will exhibit this alignment strongly toward the bottom of the treble staff. Vowels with a higher f_{R1} will find corresponding alignment toward the top of the treble staff. It is important to know that these approaches all feature a strong and pure percept, but that they sound different based on the tone color of their first formant (F_1). This is a reminder that the tone color of a strong first resonance (f_{R1}) that is aligned with the fundamental ($1f_o$) moves through |~u| and |~o| as it ascends the treble staff, followed in its ascent by |~ɔ|, |~ɑ|, and |~a| in the octave between F5 (~698 Hz) and F6 (~1397 Hz). Above that, the fundamental of any sung sound is both extremely simple and sounds significantly |~æ| in tone color. Technically it is possible to sing $1f_o$ into the |~e| and |~i| ASTC ranges, but that is just below and in the seventh octave. These tone colors do not exclusively form the overall vowel percept at any of these pitches. But they will be significant contributors. Note too that even if the first resonance of an [a], for example, is aligned with $2f_o$ when singing the pitch F4, the tone color of that now quieter $1f_o$ remains a less pronounced |~u|.

There is no one accepted practice for singing the low f_{R1} vowels toward the top of the treble staff or the high f_{R1} vowels toward the bottom of the treble staff. Much remains up to the aesthetic targets of the singer in these cases. I have found that the need to meticulously *tune* resonances based on pitch is obviated once issues related to breathing and airflow, vocal fold adduction, and basic vocal tract shape are addressed. This resonance strategy is, like the others I will cover below, an emergent property of loud, efficient singing. It is not a quick fix that solves other technical challenges.

In figure 10.6 you will find two different examples of *whoop* or *hoot*, which is the abovementioned alignment of the slowest vocal tract resonance (f_{R1}) with the fundamental ($1f_o$) to generate a high-intensity, perceptually pure F_1 . Figure 10.6 shows waveforms and spectra of a soprano singing an [a] at the pitch C6 (top), and a countertenor singing

an [i] at the pitch E4 (bottom). Note that for each pitch period, both waveforms feature a single slow ripple from low to high pressure, and back again to low. Three periods of each sample are shown, and the approximate start of the second and third period is marked with a dashed line. Note the difference in timescale as the C6 has a significantly shorter pitch period.

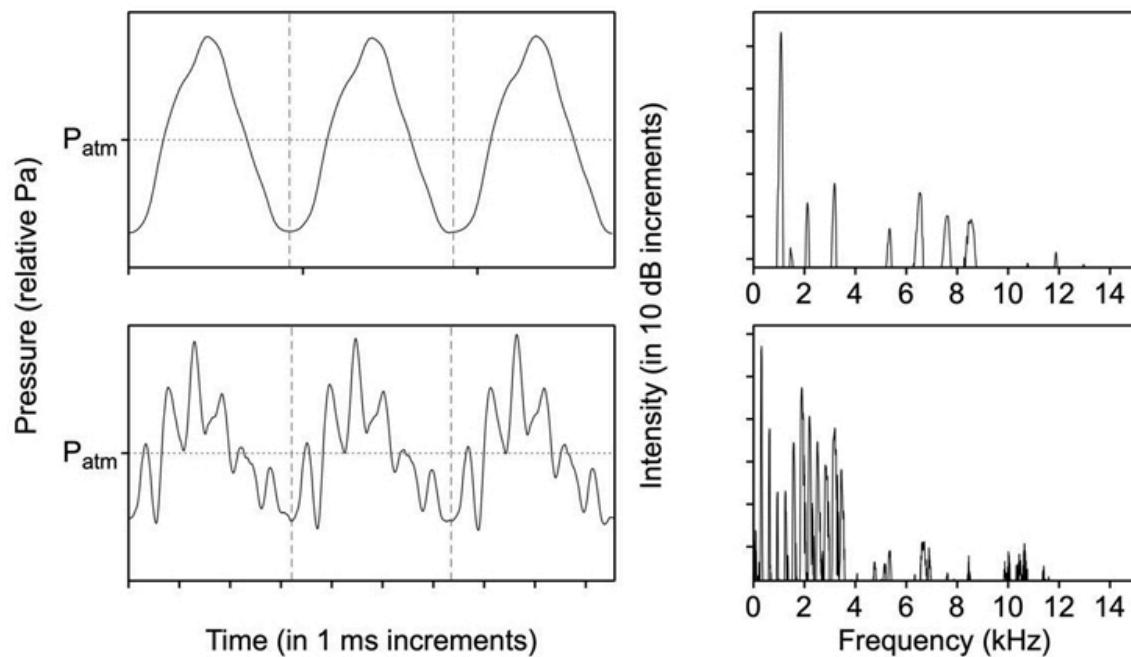


Figure 10.6 Waveforms (left) and spectra (right) of C6 sung by a classical soprano (top) and E4 (bottom) sung by a classical countertenor (note different timescale). Vertical lines in waveforms indicate estimated start of each period. Source: Recorded by author under controlled conditions.

We can start to predict the perceptual differences between these two sounds by comparing the waveforms and spectra. Both spectra in [figure 10.6](#) are dominated by $1f_0$. This is how the Fourier transform displays that slowest ripple in the waveform in a spectrum or spectrogram, regardless of the frequency. However, the soprano's waveform and spectrum are both visibly simpler than the countertenor's waveform and spectrum. The countertenor sample has additional, faster resonances

oscillating per glottal cycle. These appear as faster ripples per period in the waveform and higher frequency harmonics in the spectrum. This is a function of the vowel, the pitch, and the likely greater contacting area of the vocal folds. The vowel [i] has a higher second formant (F_2) than [a], which explains the faster ripple. The supraglottal impulse at a lower pitch has the potential for a more asymmetrical drop in pressure, which introduces faster rates of change in the supraglottal impulse. This is facilitated by a looser vocal fold cover that likely makes deeper contact.

In terms of the *sound* of this *whoop* or *hoot* strategy, the $1f_o$ of the soprano sounds between |~a| and |~a| and the countertenor's $1f_o$ sounds clearly like |~u|. The same strategy sounds very different at different pitches because tone color is tied to frequency. The countertenor's $1f_o$ elicits a tone color that does not exist in the soprano's sound because her rate of supraglottal impulses is too rapid. Remember, some aspects of timbre have duration. Because the human hearing mechanism is more sensitive to frequencies in the range of the second formant for [i] than the first formant, the strong oscillation of F_1 (|~u|) is generally not *heard* as the strongest component of this pressure pattern. The strongest perceptual experience is |~i|. Remember this for later; sometimes the most audible portion of the spectrum points us away from the strongest acoustic forces in the vocal tract.

The most noticeable difference between these two kinds of sounds may be the lack of much auditory roughness in the percept of the soprano's C6 and the presence of it in the percept of the countertenor's E4. This is not to do with those voice types so much as the pitches they are singing. In many approaches to classical treble singing, one would anticipate that auditory roughness would gradually leave the sound as one ascends from the bottom of the treble staff to pitches in the sixth octave. Even if the goal remains to have a bright sound, that brightness will increasingly lose buzziness as a function of ascending pitch.

The Tone Color of a Strong $3f_o$ or $2f_o$: Money Notes across Genres

Another common strategy to build power in a higher singing range aligns multiple repetitions of a vocal tract resonance (either f_{R1} or f_{R2}) with the pitch period. Singers across genres have worked out how to shape their vocal tract to minimize the damping of this resonance within that period. At a high enough pitch, this generates a continuous oscillation that visibly defines the pitch period in the waveform. In the spectrum, this translates to a single, high-intensity formant in the lower part of the spectrum, which elicits a strong pure percept (no auditory roughness), and it may be dull, cutting, or even edgy, depending on the pitch and vowel shape. The tone color of this single formant is characterized by the ASTC of the peak harmonic, regardless of whether that formant was generated by the first or second resonance (f_{R1} or f_{R2}). Think about that for a moment. If a vowel is strongly characterized by the presence of one dominant tone color, but fully realized by the presence of co-presenting complementary tone colors, this means that a sung sound with just one of two vowel formants present will necessarily fall short of the timbral complexity of a speech vowel. This has implications for our percept of such sounds. Also, universally present in this way of singing is higher frequency energy that is buzzy (with auditory roughness). This is common in high-energy, typically high-pitched “money” notes, and where in the spectrum this buzziness falls has much to do with genre and performance context.

Think about how this points to the importance of the relationship between the timing of vocal fold oscillations and vocal tract responses. Also, keep in mind that the voice is not a brass instrument. Unlike the resonances of a metal tube, the resonances of the vocal tract are typically not so strong that they fight vocal fold oscillation when mistimed. At these high-pressure levels though, the nonlinearities of the system do kick in. When well-aligned, these kinds of sounds are incredibly powerful on the outside and experientially easy to produce on the inside. When misaligned, the mismatch between the singer’s target loudness and the actual loudness typically results in unbalanced, effortful phonation. In a speech-pitch range, the voice is characterized by timbral options. Shade a vowel this way or that way, and you move from one regional accent to

another. Similarly shade that vowel while sustaining a high-intensity “money note,” and the entire system may stop working.

Remember, we cannot mistake the map for the terrain. I will *never* suggest that singers executing this resonance strategy think about it in these terms, or that they would need to find these solutions using a computer. Everything I know about singing tells me that singers need (1) a general sense of the target sound, vocal tract shape, and anticipated effort level; (2) an emotional charge that organizes the entire system; and (3) time to play and experiment until they find the solution that feels and sounds the best. I suspect that from the singer’s point of view, they make a relatively small adjustment as pitch rises that allow them to *keep singing*. Understanding the underlying physical reality of this adjustment is for the teacher so they can make helpful predictions in the studio.

Back to the map: The fact that we hear such a formant peak as close to the tone color percept of a pure tone means that this resonance oscillates to the end of the pitch period. The fact that it has no auditory roughness means it oscillates no more than five times per pitch period. We see these oscillations as a strong $2f_o$ or $3f_o$ (or possibly $4f_o$ or $5f_o$) in a spectrum or spectrogram. This is common in classical tenors, baritones, and belters who sing high-energy pitches between ~G4 and C5. There are also examples of this dynamic commonly found in belting and mix-belting approaches above C5 in contemporary and some musical theater sounds.

As a priming example, here are images from two samples recorded under controlled conditions. Figure 10.7 (top) shows three periods of a classical tenor employing a $f_{R2} \approx 3f_o$ resonance strategy on a C5.¹² This means that the second resonance (f_{R2}) of the vocal tract oscillates three times for every glottal contacting event, and that the final drop in pressure of the resonance occurs as the vocal folds facilitate a drop in supraglottal pressure. Note that the waveform does show a strong ripple that oscillates three times per pitch period, which corresponds to a strong $3f_o$ in the spectrum. The first resonance (f_{R1}) is significantly damped, and its effect is diminished in both the waveform and the spectrum. The second example (figure 10.7, bottom) shows a musical theater *belt* on a

G4 where $f_{R1} \approx 2f_o$. Note that the waveform shows a strong ripple that oscillates two times per pitch period, which corresponds to a strong $2f_o$. This indicates the interaction of the first resonance (f_{R1}) aligning every other oscillation with the glottal contacting events. I hope your intuition is developing to the point that you would expect a strong ripple that oscillates four or five times per pitch period to appear as a strong $4f_o$ or $5f_o$, respectively. In all these cases, these high-amplitude, continuously re-excited resonances would collaborate (time well) with the supraglottal impulse to generate a strong excitation of the vocal tract per period.

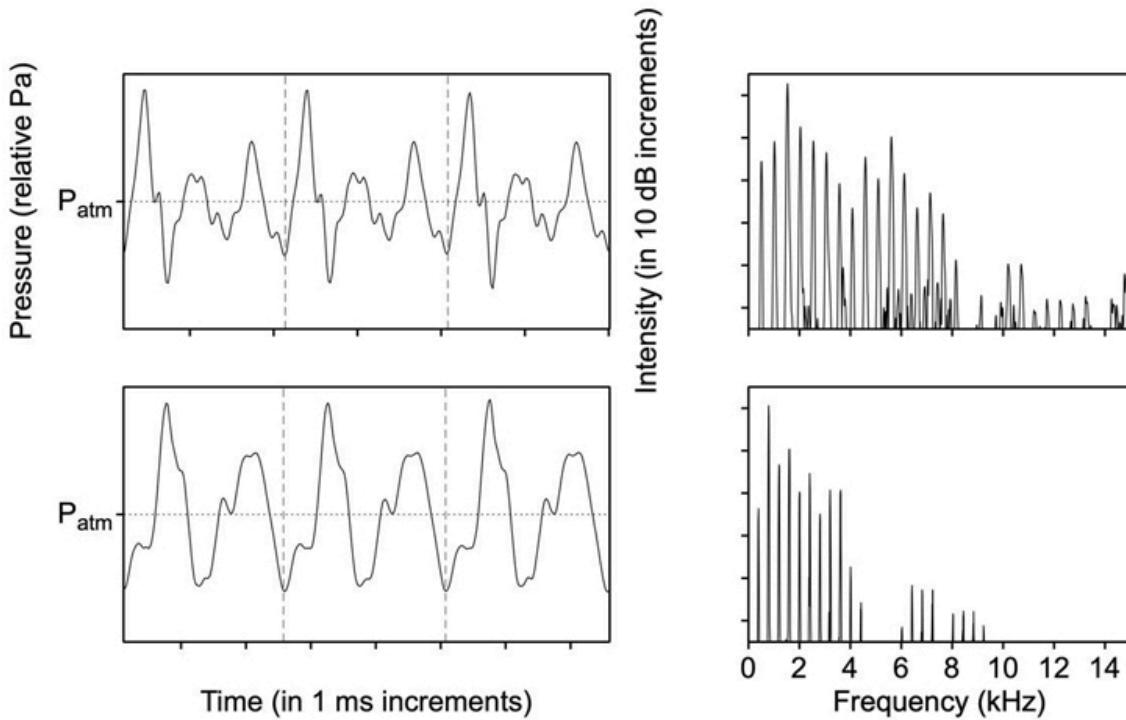


Figure 10.7 Waveforms (left) and spectra (right) of C5 (top) $f_{R2} \approx 3f_o$ and waveform and spectrum of G4 (bottom) $f_{R1} \approx 2f_o$ sung by two different tenors. Approximate start of each period marked with a dashed gray line. Source: Recorded under controlled conditions by Dr. Mark Tempesta at the Voice Pedagogy Lab at the University of North Texas and by Dr. Nicholas Perna at the Berton Coffin Voice Lab at the University of Colorado Boulder.

At this point, I need to acknowledge a limitation on the conclusions we may draw from the samples that follow, because they are taken from

commercial recordings. In fact, I include the controlled recordings above to point to the credibility of the limited claims that follow. The use of commercial recordings introduces several challenges that impact both the waveform and spectrum. One does not know the distance to or frequency response of the microphone used, nor the electronic post-processing that may have been applied. And the frequency limitations of the recording media can be severe for older recordings. Additionally, frequency filtering and artificial intelligence processing were carried out to remove the confounding effect of the instrumental accompaniment in these samples. So I am not confident that every part of the resulting waveforms and spectra display what would have been captured from these artists with a reference microphone in a controlled environment. We will never know, for example, if a given $1f_0$ would have been 4 dB more intense under laboratory conditions, or whether the loudness of the highest frequency energy is a result of either equalizing the spectrum or applying dynamic compression.

However, having worked live with many singers who sing in this way, I am confident that the ways in which these voices are captured here does a good job of accurately representing their gross resonance strategy. This remains true even if the relative intensities of their harmonics, or indeed the objective intensity of their entire voice, are unknowable. I would not use these samples to make more detailed claims beyond the general shapes of the spectra. Discount the accuracy of the waveforms and spectra of these commercial recordings if you must. Or you may strike a middle ground and hedge that we can certainly talk about what the *recordings* sound like, even if we do not know for sure exactly what the singers sounded like.

Given that we observe the same formant structures in singers recorded in a controlled environment as is revealed from the commercial recordings that follow, and given that the target vowels sound comparable in the controlled and commercial recordings, the commercial recording process for the samples that follow would appear not to have so severely reshaped the spectrum of these singers that they provide no hint of these singers' gross resonance strategies. I might go so far as to suggest that it

strains credulity to imagine that these disconnected commercial recording processes all stumbled onto popular and identifiable resonance strategies when the singer sounded radically different. I would wholeheartedly support the initiative to capture laboratory recordings of elite singers and compile a public research database. With that caveat said in hopes of assuaging the most skeptical among us, I will proceed.

Here are examples of three different singers operating in different genres, all executing a $f_{R2} \approx 3f_o$ strategy (see [figure 10.8](#)). Sarah Bareilles (top) sings a C#5 in a cutting mix-belt; Franco Corelli (middle) sings a similarly cutting C5 in his full, operatic voice, and Thomas Hampson (bottom) sings a G4. The waveforms each show three periods, and the approximate start of the second and third periods is marked with a dashed line. Note that all three waveforms show a strong ripple that oscillates three times per pitch period, which corresponds to a strong $3f_o$ in all three spectra. The ASTC of that $3f_o$ for Bareilles and Corelli—approximately 1,662 Hz (around G#6) and 1,569 Hz (around G6), respectively—falls into the |~æ| tone color range. And like the controlled sample in [figure 10.7](#) (top), both of these vowels have a bright, [æ]-like quality to them. This, despite the [o] and [a] on the page in these two excerpts, respectively.

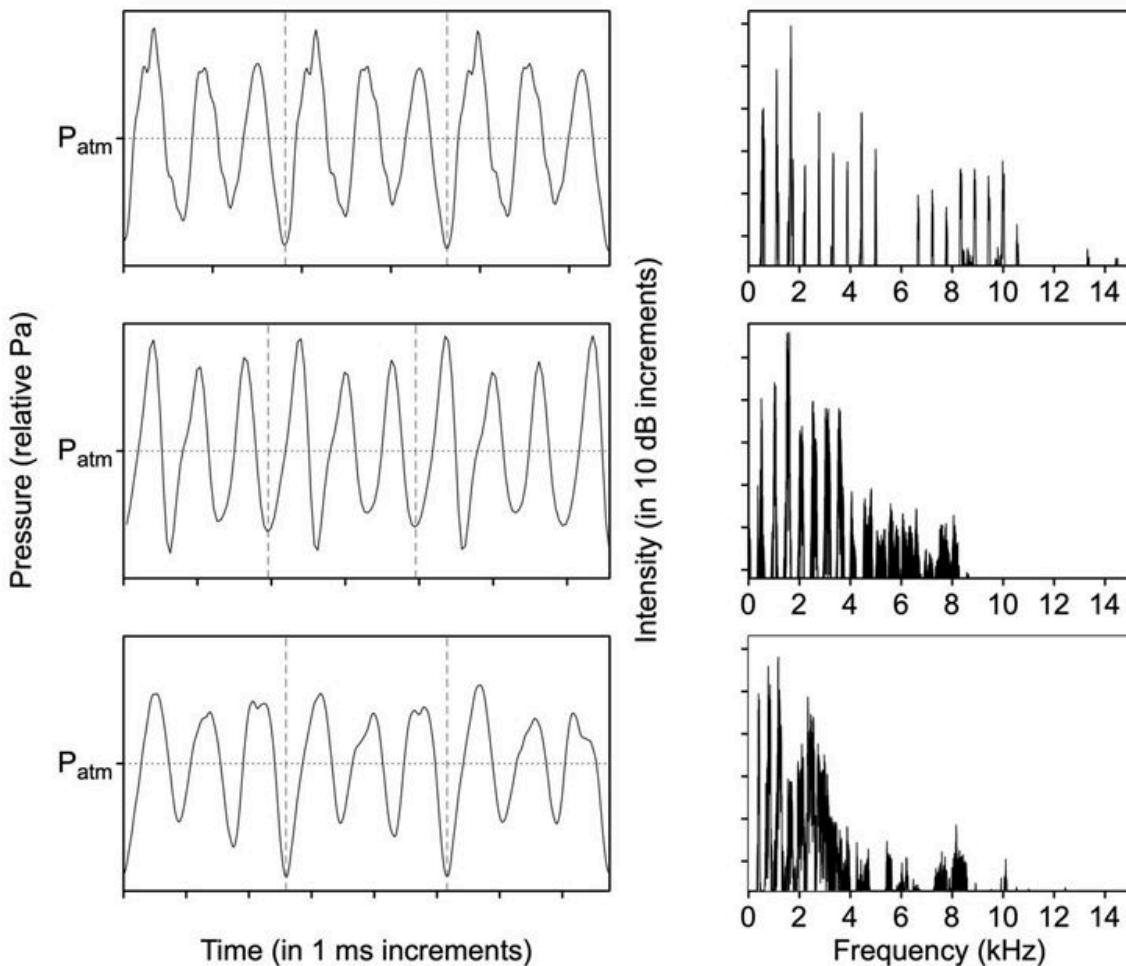


Figure 10.8 Waveforms (left) and spectra (right) of C#5 (top) from Sarah Bareilles “Goodbye Yellow Brick Road” by Elton John, C5 (middle) excerpt from Franco Corelli “A te, o cara” from Bellini’s *I Puritani*, and G4 (bottom) from Thomas Hampson “Avant de quitter” from Gounod’s *Faust*. Vertical lines in waveforms indicate estimated start of each period. Source: (Bareilles) <https://youtu.be/8nJTyja5u4g?si=IBoLJH0mf-tGMxfp>; (Corelli) <https://www.youtube.com/watch?v=jh7t0ykwPqM>; (Hampson) <https://www.youtube.com/watch?v=jh7t0ykwPqM>. Filtered by lalal.ai and by author in Praat.

The third example (figure 10.8, bottom) further demonstrates how this $3f_0$ strategy generalizes. Here we see a climactic G4 on the word ‘ma’ sung by Thomas Hampson. As with Bareilles and Corelli, he has likely aligned every third oscillation of the second vocal tract resonance (f_{R2})

with the pitch period. Also, as with Bareilles and Corelli, this resonance does not significantly dampen within the period. Does this also sound like [æ]? It does not. Because his pitch is lower, the ASTC of $3f_0$ ($\sim 1,176$ Hz) falls on the border of |~a| and |~a| rather than |~æ|. Listening to this sample, you will notice that it sounds as though the indicated [a] vowel neutralizes as the vowel formant region of the spectrum reduces to a single peak. In my experience, baritones identify this as a modification to [ʌ]. However, as with the Bareilles and Corelli examples, the sound falls short compared to the complexity of a speech-pitch vowel. And the transition from [a] to [ʌ] at lower pitches tends to decrease the overall amplitude of the pressure wave, which cannot be said for Mr. Hampson here. It may not be that he is consciously neutralizing his sound in the sense of changing vowels; it may be that the acoustic process of two formants simplifying to one brings about a loss of complexity that *sounds* like a neutralization.

This “vowel simplification” raises a semantic question. Despite the high likelihood that the second resonance (f_{R2}) has aligned with the third harmonic ($3f_0$), the damping of the first resonance means that second resonance (f_{R2}) creates the lowest formant in the spectrum. This is technically the first formant (F_1). If we view this phenomenon through the source-filter theory, this is technically a first formant (F_1) strategy.

While the above strategy favors aligning every third oscillation of the second vocal tract resonance with the supraglottal impulse, it is also common to align every other oscillation of the first vocal tract resonance with that glottal pattern. We see this referred to as an alignment of f_{R1} with $2f_0$. Donald Miller explicitly defines *belting* according to this acoustic relationship.¹³ Figure 10.9 shows two examples of this at two different pitches. Steven Pasquale (top) is shown belting a high-energy G4, and Ethel Merman (bottom) is shown belting a C5. Three periods are shown in the waveform, and the approximate start of the second and third periods is marked with a dashed line. Unlike the soprano and countertenor in figure 10.6 or the operatic and contemporary singers in figure 10.8, both Pasquale and Merman have a strong ripple that oscillates twice per

period. This appears in the spectrum as the slowest vocal tract resonance aligning with $2f_0$. Because these two samples are sung at different pitches, the tone colors of those first formant (F_1) peaks are not quite the same. The second harmonic of G4 falls in the $|\sim\circ|$ range, while the second harmonic of C5 falls in the $|\sim\alpha|$ range. Listen to just the first formant (F_1) from these two samples back-to-back and you will note that the tone color *opens* from the Pasquale to the Merman.

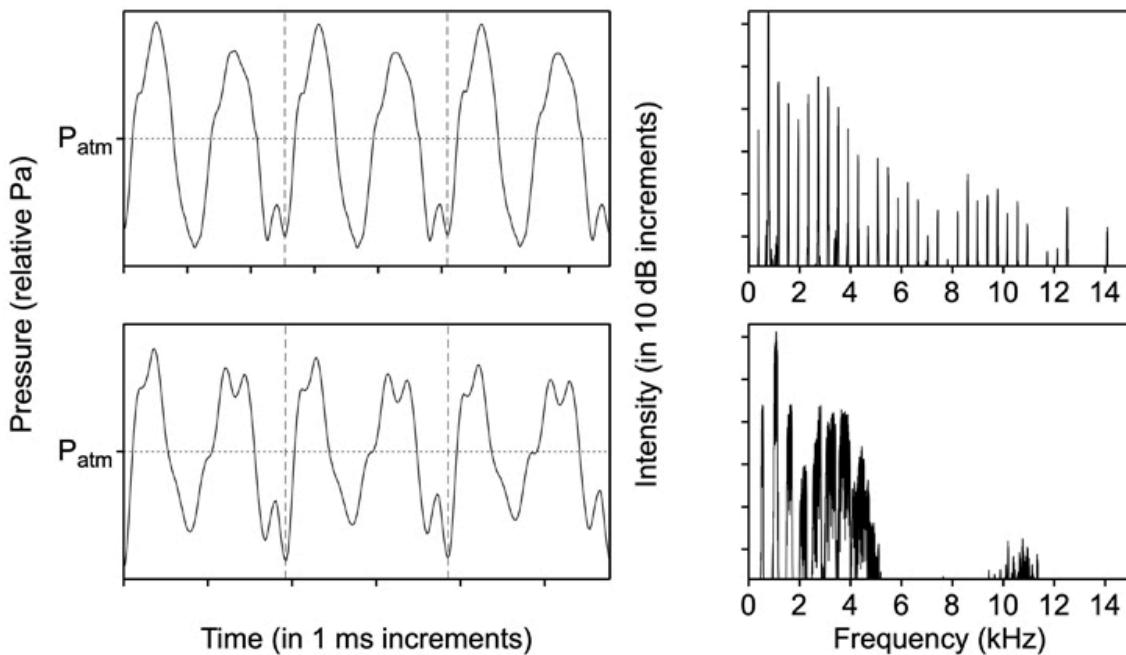


Figure 10.9 Waveforms (left) and spectra (right) of G4 (top) sung by Steven Pasquale in “Wondering” from Bridges of Madison County by Jason Robert Brown and of C5 (bottom) sung by Ethel Merman in “Everything’s Coming up Roses” from Laurent, Styne, and Sondheim’s Gypsy. Vertical lines in waveforms indicate estimated start of each period. Source: (Pasquale) https://youtu.be/voZPsEnRDUQ?si=X1SH8OQr4azEr08_; (Merman) <https://youtu.be/mPPAtQ9FpMY?si=kqpVA2hu95LvhTpY>. Filtered by lalal.ai and by author in Praat.

See Lab #26 to explore the above $1f_0$, $2f_0$, and $3f_0$ samples.

Hearing the Entire Spectrum

Now that we have considered various strategies for keeping a strong resonance oscillating within the pitch period and considered what the tone color contribution of the first formant (F_1) is in each case, we will consider the rest of the spectrum. Be prepared to apply your visual annotation skills.

Looking back at figure 10.6, we would predict that the soprano (top) would trigger little auditory roughness. She would convey a shimmery, noisy quality in addition to the dominant ASTC of $1f_o$. All but two of her harmonics shown here fall within the $|\sim i|$ or $|\text{bright } \sim i|$ tone color ranges. Her top four harmonics fall above the 5 kHz pitch perception threshold. She would have a fundamentally uncomplicated timbre characterized by the fundamental and the bright and buzzy “rest of the spectrum,” which you may recall we characterized earlier as sounding like crickets. The countertenor, by contrast, has a lot of energy in very bright, rough, and noisy spectral regions. His timbre would feature a strong $|\sim u|$ from the fundamental, and at least two qualities that sound like darker and brighter versions of $|\sim i|$. Or more likely, those qualities blend into a wash of $|\sim i|$. His second formant (F_2) is buzzy, and the next formant above that may not resolve into the pitch. Additionally, there is some higher frequency noise in the spectrum, although it is of very low intensity.

Looking back at figure 10.8, let us compare Bareilles (top) to Corelli (middle). Again, I acknowledge that the recording process has certainly reshaped the high-frequency parts of the spectrum here. But notice the difference in harmonics over 5 kHz, and that the energy clustered around 3 kHz in Corelli’s example is more broadly spread between about 2.4 kHz and 5 kHz in Bareilles’s. Comparing these two recordings, we can note that Bareilles has much more noisy brightness than Corelli does in this example and more roughness across a wider range of $|\sim i|$ tone colors. Corelli’s voice will likely sound very powerful, but in a more timbrally compact manner. Thomas Hampson’s spectrum (figure 10.8, bottom) looks remarkably similar to Corelli’s, as though the entire formant structure were transposed down a fourth. Because those formants slowed down for Hampson, the tone color changes.

Looking back at figure 10.9, I am struck by the delightful similarities between Pasquale (top) and Merman (bottom). As with Hampson compared to Corelli, Pasquale has transposed the formant structure down a fourth compared to Merman. With the exception of additional noisy high-frequency harmonics in this recording, Merman, Hampson, Corelli, and Pasquale all have managed their singer's formant portion of the spectrum in a similar manner. Unlike Bareilles, who leverages noisy brightness across a wide frequency range, most of the bright energy of the other subjects concentrates between 2.4 kHz and 3 kHz.

I will unpack some of the implications of the above in the next chapter, but I would like to make one observation now. In every one of the above examples, regardless of genre or pitch range, these singers manage to sing in a way that keeps a high-amplitude oscillation going in their vocal tract. Because the vocal tract is a more favorable resonator to lower frequency resonances, each of these singers has leveraged a resonance that aligns with the next supraglottal impulse after only one, two, or three oscillations. Much faster resonances are also excited and decay per pitch period; indeed, bright sounds with an |~i| tone color or brighter are a strong feature of every one of these samples except the soprano singing a C6. However, those faster resonances are clearly lower in amplitude in the waveforms. In a physical sense, the slower resonances are significantly stronger oscillations.

Once again, we return to one of the many paradoxes of understanding both how we produce and how we perceive voices. The ear is more sensitive to pressure oscillations between about 1 kHz and 3.5 kHz than slower or faster oscillations. This means that we may hear the bright part of the sound as louder, even though the warm part of the sound is helping drive phonation. This is especially confounding at a distance, when the ear becomes even less sensitive to lower and higher frequency stimuli, selectively preferring the midrange. So while it is true that many ways of singing favor a strong, bright |~i| quality, setting that brightness as the initial training target may cause the warmer parts of the spectrum to suffer. It is those harder to hear, warmer percepts that the singer needs to sort out first to secure their technique, because they literally signify that

the system is working efficiently. Put another way, I would much rather start with a singer who has stabilized their voice in a way that would welcome introducing brightness to the existing warmth than try to undo the effort required for a bright sound generated by an excess of tracheal pressure and physical effort.

WHAT CAN BE PERCEIVED?

Surely experts hear with more specificity than novices do. It is impossible to guess how much of what we have explored can be heard by the naive listener, especially on a first pass. Sustained tones permit deeper critical listening. Obvious contrasts are the easiest to discern: buzzy versus pure, and noise versus pitch. Even this basic approach to listening can develop one's ability to hear functionally. Revisit the two-timbre model in [chapter 7](#) to consider how changes at the glottal level rebalance the amount of auditory roughness and pitch resolution in the sound. The more dissimilar two aspects of a spectrum are, the easier they are to distinguish. This is especially so the further apart they are in tone color.

In pitch ranges where the lowest two resonances both generate a strong, discrete radiated peak, the overall buzziness of the percept may prevent the separation of those tone colors beyond simply noticing their relative warmth and clarity. As pitch rises and auditory roughness falls away from the vowel-defining portion of the spectrum, more specific tone color qualities may become clearer. This is evident for vowels such as [e] and [i] because the tone colors of their f_{R1} and f_{R2} are less similar than the f_{R1} and f_{R2} in vowels like [o] and [a].

As we consider how to find these resonances by their sounds, we must be cautious in attempts to elicit the characteristic color of a specific vowel's formant by broadly exciting the vocal tract. Scott McCoy (2019), among many others, suggests flicking the throat while the glottis is held closed. Doing this while changing vocal tract shapes for various vowels will play a little melody.¹⁴ It is actually a great way to demonstrate that resonances are time-and-pressure phenomena. We hear F_1 as a pitch in this demonstration specifically because the resonance periodically repeats in the vocal tract at a frequency that triggers a pitch percept. However, the tone color produced by flicking the throat will be the combination of all the resonances of the vocal tract sounding at once, not exclusively the tone color of F_1 . If you record yourself doing this and pass-filter just the F_1 response of the vocal tract, you may be surprised that other tone

colors exist in the higher formants. McCoy (among others) similarly suggests whispering to hear the pitch of the second formant (F_2). The same caveat applies—that the tone color will be dominated by, but not exclusively be, that of the second formant.

Other methods to excite the vocal tract to reveal its approximate formant frequencies include vocal fry, white noise from an external speaker, or an externally generated impulse (either from a recording or by flicking the throat). We must exercise caution not to extrapolate these sounds to a pitched sound. Pitched phonation not only typically features no spectral information below the fundamental but also reflects a completely different excitation pattern for the inner ear. These vocal tract excitation exercises also contain no helpful information regarding auditory roughness or pitch resolution, and they do not capture tone color changes that occur because of the pitch being sung. See Lab #27.

Singing above C4, the presence or absence of $1f_o$ may be clearly heard. In the case of a classical treble singer, this will likely correlate to the timbral complexity of the voice, or its maximum potential amplitude. Lighter *Fachs* tend to have less power in this regard. One also hears the characteristic tone color of the fundamental in a variety of contemporary singing styles. As mentioned above, this is clearly audible in the opening minute of Whitney Houston's recording of "I Will Always Love You. Another good example is the first 0:48 of Nina Simone's "Feeling Good."¹⁵

Unbalanced classical treble voices can get into trouble when they do not take advantage of this power. This typically sounds like strong bright energy with little power in $1f_o$. These singers may depend instead on pressed phonation or tracheal pressure well above the phonation threshold pressure. One may hear issues related to intonation, legato line, or a general lack of easefulness that accompany an otherwise exciting sound. This is to their long-term detriment, despite how thrilling a sound it is. In a classical tenor or baritone, musical theater, or contemporary singer, the absence of a strong $1f_o$ likely reflects a brighter tonal goal and lower tessitura. Singers using directional microphones at a distance that triggers the proximity effect (typically within 30 cm or 1 ft.) will enjoy a boost to the warmest portion of their spectrum. This must be kept in mind

when training without a microphone, and when selecting and positioning a microphone for an online lesson.

CONCLUSIONS

I hope it is sinking in that as complex a perceptual experience as the voice is, one may begin to parse that sound out into discrete and perceptually opposable chunks. Is it dark or bright, is it pure or buzzy, is the vowel complex or simple, is there high-frequency noise or not? These are remarkably simple questions to ask, and with the help of a spectrograph (for example VoceVista Video Pro or Praat), one may quickly learn to identify these qualities. The final chapters will attempt to assemble in one place the best information I have as to the functional correlates to these various sounds. It is my sincere hope that you have listened to the samples and explored these labs along the way, and that you are ready to make new connections between what a singer does and what sounds they make.

NOTES

1. At very low and very high pitches, this rule of thumb becomes less accurate. The equivalent rectangular bandwidth ([chapter 7](#)) should be used to explore these limitations.
2. Review [chapter 5](#) for a detailed account of pitch resolution.
3. Kenneth Bozeman, *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013), 26–27.
4. Steven Austin, "Provenance: The voce chiusa," in *Journal of Singing* 61, no. 4 (March/April 2025): 421–26. Austin offers a lucid discussion of these terms as they arose in Manuel García II's work.
5. Bozeman, *Practical Vocal Acoustics*, 23–24.
6. Donald G. Miller, *Resonance in Singing: Voice Building through Acoustic Feedback* (Princeton, NJ: Inside View Press, 2008), 52–53.
7. Bozeman, *Practical Vocal Acoustics*, 23; Miller, *Resonance in Singing*, 52.
8. Richard Miller, *Solutions for Singers: Tools for Performers and Teachers* (Oxford: Oxford University Press, 2004), 75.
9. Richard Miller, *Training Soprano Voices* (New York: Oxford University Press, 2000), 117.
10. Ian Howell, "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color" (DMA diss., New England Conservatory of Music, 2016), 41–43.
11. Whitney Houston, "Whitney Houston—'I Will Always Love You' (Official 4K Video)," [4:34], YouTube, <https://youtu.be/3JWtaaS7LdU?si=UjKZOM-iZYg28IiB>.
12. Miller, *Resonance*, 3.
13. Miller, *Resonance*, 111.
14. Scott McCoy, *Your Voice: An Inside View*, 3rd ed. (Delaware, OH: Inside View Press, 2019), 68.
15. Nina Simone, "Feeling Good (Official Video)," YouTube video, 3:56, published by Verve Records, July 24, 2013, <https://youtu.be/oHRNrgDIJfo>.

11

How to Teach Singing

This chapter describes how I both teach and coach singers, and how I mentor others to do the same. Woven throughout are practical frameworks that will begin to inform how *functional listening* plays out in various contexts. In some ways the previous chapters were easier to write than this one because logical arguments based on and extended from a set of facts can be methodically structured. Voice teaching is much less straightforward a process, rarely linear, and only occasionally susceptible to even the best-formed argument. This chapter is not directly about perception. It is about how to incorporate perception into a cohesive approach to teaching.

Voice teaching leverages “what works” more than what is factual; “what is helpful” more than what is true. There are powerfully useful lies in voice teaching that cannot be rigorously tested. So this chapter may seem like a departure from the previous ones, appearing at times like a series of generalizable concepts that I have found to be effective.

For the singer, singing should feel like a unified, expressive response to a simple thought. A well-coordinated, artistic singer does not use knowledge of facts to correctly position their body in space and drive vocal fold oscillation with the right amount of tracheal pressure. That optimal coordination does not then enable

the singer's musical intuition to use that technically efficient body expressively. This is exactly backwards. We must get the singer's *thinking* right first, which will stimulate a response in their body. Once the singer falls into that pattern of stimulus (the thoughts) and response (the singing) without additional, counterproductive steps of self-monitoring and attempts to independently control separate parts of the body, we can *start* the work.

The flip side of this is also true though. The teacher who wants to truly understand *how* to teach effectively is arguably best served by deeply understanding the entire process rather than a personal, procedural narrative. My tongue is perhaps a little in my cheek as I write this, but I frequently find that those who argue for the immediate practical application of a new idea—or its dismissal—do so to make *their* learning process easier, not that of their students. Ideally, the role of the teacher or coach is to guide the singer through a series of experiences based on, informed by, and shaped by the logically formed frameworks within the teacher's mind.

If you have read this book (or parts of it) in a voice pedagogy class, look inward as you contemplate the material. Is your goal to exclusively operationalize the conclusions of others, or are you here to transform the way you imagine how the voice works and how you can work the voice? Are you limited by the solutions your teacher has observed in the genres of music they teach? I think it is not enough to know what *should* work in each circumstance—the *correction* for a given *problem*. The effective voice teacher is the teacher who has the tools to think on their feet and react productively to the dynamic environment of a voice lesson. This requires an active mind to start, one capable of hearing and effectively evaluating the functional reality of the singer.

I believe the purpose of a voice lesson for the singer is to practice the process of singing. The outcome of that lesson should be that through singing, they will come to understand how they can sing

new things in new ways. The language of teaching is whatever allows the singer to sing better. I would be happy if one of my voice teaching mentees never spoke any of the words in this book in a lesson. The ideas in this book are for the teacher to inform, to suggest, to predict, and to serve as guardrails.

With that said, in the following chapter (1) I will lay out a short primer on how to train a singer to think, and how you as a teacher can reason through where in that thought process the singer is getting caught up. (2) I will lay out several practical concepts to keep in mind as you teach. (3) I will explain how I imagine the singing body works as a functional whole. This part of this chapter will draw on the information presented in multiple previous chapters, and I aim to synthesize these ideas into a single, perception-focused narrative to understand voice production. There are many parts of the singing body to be sure, but there are ways to conceive of the act of singing that take us out of the provincialism of anatomical and acoustical subsystems into a global view of how that entire system can work together. Understand those interactions, and you can start to understand how different ways of using the body physically produce different sounds. You can then start to think about intermediate sounds and adjacent functions that fall short of the target sound. But in falling short, they establish some crucial functional coordination that allows the singer to quickly reach that target sound with a procedural understanding of how they did it. (4) I will present three good rules of thumb that connect perception to vocal function.

HOW A SINGER SHOULD THINK

The simplest, but most important pattern you need to know about is the dynamic of stimulus and response in the singer.¹ In this case, stimulus means what you *think* as you ask your body to sing. The response is the way in which your body *responds to that stimulus*. This includes the myriad patterns of muscular engagement, the effects of postural effort and one's sense of proprioception, the efficiency of breathing and phonation, and the way phonation and resonation interact. In efficient singing, these do not feel like separate concerns. Rather, they are aspects of the functional whole of the singer's response. To be sure, I want to be cautious not to imply that the brain is segregated from the body. But we can be complete beings while recognizing our ability to consciously decide to move our bodies in certain ways.

Most singers who seek lessons have not yet clarified the stimulus, nor have they found the efficient, spontaneous response they desire. As such, *your* exploration of *their* stimulus-and-response patterns will necessarily include breaking down some aspects of that ultimate, seamless, all-at-once act into smaller challenges. Our goal in discussing the parts of the whole is not so that we can focus on the parts. For rapid improvement, we do well to understand the coordination of the *whole* in terms of actionable choices that target specific issues in specific aspects of the singer's process.

THE FIVE STEPS IN ALL SINGING

I would like to explore how one can structure their thinking while training their voice. This framework came from my teacher, Lynne Vardaman, a student of Dan Merriman and Cornelius Reid in New York City. She taught me that as problematic as overthinking can be for a singer, the answer is not to “avoid” thinking, or “ignore” those thoughts. Her answer was to think specific thoughts at specific times in the process of singing. The mind does not wander when you give it something to do.

What is the thought process of singing? Vardaman suggested breaking singing down into discrete steps that take place in repeating a sequence. These steps are mutually exclusive, which is to say that you cannot go back to a previous step and correct what you just did. Executing Step 1 effectively leaves your body and mind in exactly the right place to begin Step 2. Step 2 similarly hands off to Step 3 with no intermediate thought or action. This continues, and the last step connects to the first step of the next cycle. These steps are not just conceptual, they are actual words that you think. When you give your mind something specific to think, and you keep that thought simple, you may be surprised how quickly an “overthinker” can directly elicit the desired response and how quickly a “non-thinker” can deepen their process.

These steps are (1) release, (2) clarify, (3) engage, (4) continue, and (5) release. Before I explain each step, please keep in mind that this is one way to organize thoughts as a part of training. Once the response of the body is a direct result of the stimulus, the thinking becomes even simpler. Eventually you just think the thought you want to express, and your body will organize to accomplish that. Think of this as a pedagogically effective crutch to use for a short

time so that it becomes possible for you to directly interface with your student's process.

Release refers to how you breathe in. I will go into more detail about breathing later in this chapter, but I do not advocate any dogmas related to respiration. For now, imagine that actively thinking the word *release* encourages several general responses in the body. First, you do not engage any thoracic muscles you would then need to release to continue inhaling to whatever your target is. This is a somewhat common response pattern in uncoordinated singers. It especially reveals itself in a rigid abdominal wall that paradoxically engages and releases in an alternating pattern during the inhalation. Similarly, some singers will attempt to use only muscular effort to suspend or expand the rib cage at the beginning of the inhalation rather than allowing its expansion in response to the displacement of air within the lungs. This typically stiffens the rib cage and inhibits its free reaction to the activity of the rest of the thorax. Different ways of breathing facilitate different ways of singing. However you inhale, inhale directly. The first step is to ensure that you do not meet that inhalation with restriction or effort. Hence, think "release."

Clarify suggests that you need to imagine what you are going to sing. This idea comes up in various ways in many different approaches to voice teaching. The core concept is that you must stimulate your body with the thought of the exact sound you want to make for your body to make that sound. Once the basic mechanics of a released inhalation have been established, one may immediately think this clarifying stimulus from the start. To be clear, I am not suggesting that you think about how to physically make the sound. Imagine the sound itself. You cannot start singing first and then figure out what you meant to say once the body has been organized. The mental stimulus to make the sound organizes the body.

At first this practice likely needs to be as simple as thinking the vowel sound of an exercise or vocalise. As you move from those simpler thoughts to tackling repertory, you establish a pattern of always thinking the next thing you are going to sing. In this way, the activity of singing is actually your thoughts just before your body sings those sounds. Eventually you will notice that the actual singing always drags just a fraction of a second behind your clarifying thoughts. Or, put another way, your intention to communicate is always just a fraction of a second ahead of your body's response. Your thoughts are about the future, and you never make a sound you didn't intend to make. This is a great step to slot in musical, expressive, or otherwise affective thoughts, and it takes place either during or simultaneous with the release stimulus.

Engage points to an obvious idea, but one that is worth saying out loud. When you breathe in, you are not exhaling. I know. . . that sounds foolish. But think about this as an example of the mutual exclusion of all these steps. Once you have started singing, you cannot go back and think about the sound you wanted to stimulate, nor can you go back and change the way you breathed in. Engage, then, is the moment you start to make the sound. Consider staring at your fingers and *thinking* "move," but not actually commanding the motor task. Thinking the thought and commanding the task at the same time is a different procedural dynamic in the mind and body that we can all notice. This captures the essence of "engage." It is the *doing* of the clarified thought. You might also think "go," or, in the case of an exercise or vocalise, think of the vowel itself as a verb. *Vowels are something you do, not abstract ideas.*

Shortly after thinking engage, you must think **continue**. This is the practice of continuing to stimulate the clarified sound while singing. If you take a singer through an exercise on an [a] and notice that vowel shifts to a neutralized sound almost immediately, they are not thinking "continue." Choral directors will say things like,

“refresh the vowel” to remind their singers to remember what they are doing. Your mind must keep stimulating the idea of the sound you want to make.

For a while, this part may feel ridiculous. Years ago, a student of mine described his breakthrough in understanding the step by saying, “Oh . . . I have to think [a] [a] [a] [a] [a] [a] [a] [a] [a], etc., as I sing this exercise.” Your body’s response will not continue to fall in line with your goals if you stop telling your body what to do. Conversely, you may find that many issues clear up quickly once you habituate the process of planning what you mean to do.

The final step is to **release** again, which becomes the start of the next phrase. In this way, a three-minute song is not three minutes long. It is a series of phrases likely no longer than eight to ten seconds each, at most. Every release is a complete reset of your body. What came before is irrelevant to what is about to happen. If you dependably sound worse the longer you sing, target this step.

Once you habituate this process, it is easy to observe where a singer goes off the rails. I swear, based on the sounds many singers in training make, that their clarified thought is “please don’t mess up; please don’t mess up.” Singing from a place of panic or fear will not work, as you cannot correct issues after they have already arisen. This five-step process may ultimately have a calming effect on the singer as it significantly reduces their cognitive load.

THE REGISTRATION TRIANGLE

Lynne Vardaman also introduced me to Cornelius Reid's conceptual framework that I will call his *registration triangle*, which has great utility in the clarify and continue steps. This concept neatly links functional vocal training to psychoacoustics. In short, to sing a sound with a clear stimulus and a spontaneous response, you must intend the *pitch*, the *intensity*, and the *vowel* you want to sing. Reid believed that these thoughts were the only "control factors" capable of regulating as complex and involuntary a system as the singing voice.²

Demonstrating these control factors is easy. Try to sing a high pitch (below around A4) on a quiet [u] and you will have organized your voice in a manner that finds affinity with the term *falsetto* or *head voice*. In physical terms, you are likely decreasing engagement of your thyroarytenoid muscles, stretching your vocal folds with your cricothyroid muscles, and handing over the mucosal wave of self-sustaining vocal fold oscillation from the slack cover riding over the stiff fibers of your thyroarytenoid muscles to the now stiffer vocal ligament as those muscles are stretched. Your vocal fold contacting area will decrease, especially in the top-to-bottom (superior-to-inferior) dimension, and your arytenoid cartilages will likely not make complete contact at the posterior end of the vocal folds. Additionally, you took a breath and regulated tracheal pressure to meet or exceed the phonation threshold pressure required to start that sound. The radiated sound will be dominated by the tone color of the fundamental and will have little to no auditory roughness or pitch-less noise.

Now sing a low pitch on a loud [a], and virtually every one of these functional adjustments will have reversed to align with what might be called *chest voice*. You will have shortened your

thyroarytenoid muscles, creating a loose mucosal cover and releasing the strain on the vocal ligament. Your vocal folds will contact with greater depth, and your inhalation and tracheal pressure will readjust for this sound. Your sound will likely gain auditory roughness and some pitch-less noise, and the tone color of your fundamental will be darker than your two vowel formants. Bleat like a sheep (*baaaaaaaa!*) and your tongue will move back in your mouth, further changing the acoustic properties of the laryngopharynx and pharynx to favor brightness and roughness. We can easily cause changes in the functional coordination of a singer by targeting the stimulus that results in the desired response. And we can break that stimulus down into bite-size, human-friendly ideas.

Kayla Gautreaux, Assistant Professor of Voice and Vocal Pedagogy at the Boston Conservatory at Berklee, rightly points out that while pitch is a cognitive experience, intensity and vowel point to measurable aspects of the pressure wave. If this book has consistently championed anything, it is to prize the perceptual experience of the voice over its measurable aspects. The perceptual equivalents of intensity and vowel are loudness and timbre. Therefore, I encourage us to update Reid's work to pitch, *loudness*, and *timbre*.

Reducing a sung sound to three parameters may feel too reductive. But if the goal is to sing with a free and unencumbered voice, freedom in this sense does not mean the absence of oppression; it means the presence of options. Think of Reid's framework as a reminder to explore a wide range of sounds across the continuums of pitch, loudness, and timbre.

This triangle also points to the interconnected nature of these ideas. They become discrete levers to adjust the singer's registration. The rule of thumb is that if one control factor is prescribed and you want to change a second, registration is

characterized by the way the third must accommodate. For example, if a glide from Bb3 to F4 is prescribed and you want to preserve the loudness, your timbre will have to accommodate in some way. Or if you wanted to preserve your timbre, you would have to modulate your loudness to rebalance the registration.

One of the most dependable tools I use in a voice lesson to address unbalanced registration in the middle range of the voice is to have the singer use the loud-to-quiet and open-to-close timbre paths to trigger myriad registration options in between what might feel like mutually exclusive or broken registers. The rebalancing of the system that results from reducing loudness and closing the vowel (and anticipating a loss of auditory roughness) as pitch rises can be transformative for the singer, especially when coordinating mix or other potentially nonintuitive functions. Thinking in these terms allows you to parse out registration options with great nuance.

A FEW PRACTICAL THINGS TO KEEP IN MIND

Is It Consistent? Is It Representative?

One of the changes I want to see from all my teaching mentees is to move away from them talking and toward the singer singing. There are, of course, things to talk about in a voice lesson, especially once a singer is approaching a performance. However, by and large the best way for a singer to learn how to sing is to sing. Any time you take them through an exercise, let them try it several times before you give feedback. You may even be surprised to witness the student fixing the issue on their own once they have a chance to explore it procedurally.

The second reason to let them sing is that it moves you toward the primary technical outcome of voice lessons: We want to raise the *average* performance level of the singer while narrowing the variability in outcomes between attempts. If you listen to a singer once, you have no idea whether that sample was representative of their average performance. It may have been the best or worst they have ever done. Until you have listened enough times to a singer to understand their habitual patterns of stimulus and response, intervening will not help. If their variability is wide, positive feedback may result in a worse subsequent attempt and negative feedback may appear to help. This is because no matter what, their next attempt is most likely to represent their average ability to execute the task. We don't build reliable technique around the outliers.

Different Parts of the Voice

I have found that basically every voice has three pitch ranges that demand different approaches. These are not registers or approaches

to registration. These have to do with the basic functional behavior of the system, and the boundaries of these ranges depend on voice type and genre. In a speech-pitch range, non-pathological, non-injured singing voices tend to work reasonably well when singing loud. Familiar auditory targets generally elicit a balanced response in a short amount of time, and you have many motor plans to call upon. This range tends not to *fail*.

In the highest parts of the voice, loud singing tends to *fail poorly*. This means if singing does not work well, it does so spectacularly. Balanced voices tend to respond well to the stimulus to sing high pitches. Approached with miscoordination, and the issues presented in this higher range tend to be obvious to teacher and student alike—for example, strain, cracking, flipping, high effort, or loss of power.

By contrast, the middle range of the singing voice tends to *fail well*. Miscoordination in this range may be hidden, and it does not preclude phonation. This is the most challenging range of unbalanced singing, where a singer can get away with a choice that would never work at higher pitches or be necessary at lower pitches. While the challenges related to singing extremely high pitches have tended to capture the collective imagination of voice pedagogues and voice scientists—and I frequently joke that for many of us, the term *voice pedagogy* actually means “How do I sing high notes loud?”—the subtler migrations and microadjustments to pitch, loudness, and timbre in the middle of a singer’s range tend to be where most voices are successfully balanced.

What Does a Resonance Strategy Feel Like?

We have come to one of the great paradoxes of studying the science of the singing voice to then teach people to sing. Learning how the voice works so that you can work the voice is an old and well-worn approach. Scientific research has amassed a stunning amount of

information about *what* the voice is physically and how it might possibly function. Science also offers the chance to study truly amazing singers and to analyze the acoustic output of their singing in objective terms. However, we must be cautious to not confuse the map for the terrain. Consider Franco Corelli, a singer whose recording I analyzed in the previous chapter. In that sample, he appears to have aligned his second resonance (f_{R2}) with his third harmonic ($3f_0$), and that definitely affects the sound he produced in a predictable way. Was he aware of that? Could he put his awareness of how he sang those thrilling high notes into terms that would align with a scientific analysis? As a singer myself, I assume not. The thing that can be described from the outside is frequently experienced by the singer according to a different map because, again, we must reduce the stimulus to sing to a simple idea. In some ways, effective voice teaching involves using declarative language about and demonstrations of a procedural experience in order to generate a procedural experience that the singer may then ultimately describe with their own declarative language.

The one thing I have repeatedly observed among elite singers executing difficult motor tasks is that to them, it feels good. It feels like they *keep singing* through whatever adjustment produces a cheer from those of us who can measure aspects of singing. That measurement is the outward evidence of whatever procedural solution the singer has figured out for themselves. Even something as simple as observing Corelli's mouth shape tells you little about the behavior of his tongue, which I would wager had much to do with the reduced damping in his second resonance (f_{R2}). Again, anything that can be measured and analyzed in a coordinated singer almost certainly just feels like good singing to the singer. They don't necessarily *work* for the subtle acoustical adjustments that separate

professional singers from students, nor do they experience them in a manner consistent with the output of a computer.

HOW THE ENTIRE SYSTEM WORKS

I wrote at the beginning of this book that the voice pedagogy and vocology literature tends to subdivide the singing body into subsystems. These subsystems tend to form chapters in voice pedagogy and vocology books, and we draw upon adjacent fields to present an interdisciplinary framing of the material. Conceptually, cognition and neuroscience must sit at the beginning, although this is only now entering the literature. The skeletal and muscular systems follow. Respiration is presented separately from phonation. Resonance may or may not be presented separately from acoustics. And articulation is left until the end. I do not argue against this organization, although I think we should certainly tack on cognition to the beginning, and radiation and perception to the end of this list, and situate the entire venture in the context of how we learn and communicate. However, what I do argue against is any hint that these systems are actually separable.

In this subsection, I want to introduce a model for thinking about how the singing body works. Yes, the brain is likely the most important instigator, but understanding how the physical body works as a functional whole while singing offers immense value and fertile opportunity for insight. Yes, these ideas could be expressed far more compactly with mathematical formulas, but our minds respond to narratives.

Thanks in large part to the work of Ingo Titze, you may have already encountered the idea that the voice has nonlinear properties.³ Indeed, the steps involved in phonation do not describe a linear system. Each of its subsystems needs to transfer energy (the ability to do work) to the next subsystem. The way the body does this is through asymmetries and time lag. Respiration does not drive phonation, which drives resonance, which then creates the

voice. Respiration tries to start phonation, but phonation lags a little in its response. Similarly, the excitation of the vocal tract by phonation lags the oscillation of the vocal folds. These are the asymmetries that allow energy to transfer from one subsystem to the next. In a way of thinking about it, the ability to do work is temporarily stored in the next subsystem as it is transferred, almost like storing up power in a battery, only to immediately discharge it. Because this energy spends a moment interacting with the next subsystem before being released, the properties of each subsystem can affect the nature of that release.

At the core of the concept of nonlinearity is the idea that the input of a system is not proportional to its output. A *linear* system would output one widget for every unit of energy you put into the widget machine. A *nonlinear* system does not work this way. A nonlinear system might output one widget per energy unit for some portion of the potential energy input range, but then start to output exponentially more widgets per energy unit above a certain threshold. There may even be a middle region of peak efficiency, with a weak(er) output below and above that threshold. But in the “sweet spot”—the middle of that range—the output appears to outpace the input. For example, anyone who drives an internal combustion engine car knows that fuel efficiency (your miles per gallon) peaks around highway speeds. Broadly speaking, the voice compares nicely to this kind of nonlinear system. Each subsystem does not just transfer energy to the next subsystem. It does so with a potential range of reactions, which can vary the output of the entire system and even modify the behavior of the upstream system.

As I related back in [chapter 3](#), a basic model of the singing body would describe that the pressurized air below the vocal folds causes them to oscillate. That pattern of oscillations is filtered by the vocal tract and radiated as the sound. Let’s spend a little time revisiting

that now, while turning our focus to the interactions of each subsystem and how they correlate with various perceptual qualities.

Respiration

The primary muscle of inhalation is the diaphragm. This large, domed muscle separates the abdomen from the rib cage. The contraction of the diaphragm cannot be felt well by a singer, but we can control its actions and know them by their effect on the body. There are many strategies for inhalation, including some that are counterproductive to singing. In singing, the diaphragm typically pushes down against the waterlogged viscera in the abdomen as it contracts. These organs are not compressible, and they transfer the downward motion of the diaphragm both down and out around the abdomen. Interesting things happen when this descent is arrested by abdominal resistance. This may be because of muscular engagement in the abdominal wall, or perhaps because the baseline tone of those muscles prevents further stretching. When either occurs, the breath may appear to go out to the sides, expanding the rib cage, the lower back may lose some of its curve, and the tailbone may rock forward slightly. The diaphragm is a complex structure with multiple connections that affect several different parts of the thorax both simultaneously and in sequence. Ideally its motion is smooth, and its effects are gradual.

The descent of the diaphragm appears to influence the vertical position of the larynx in the throat.⁴ The larynx drops when we yawn, for example, in part because a yawn encourages a deeper inhalation, that further contracts the diaphragm that pulls down on the lungs, trachea, and eventually the larynx. Two behaviors of the larynx in response to this lowering are important.

1. The vocal folds appear to more readily adduct when the larynx is high, and they pull slightly away from one another at baseline as the larynx lowers. You can play with this by phonating with two different lung volumes. There will be a necessary difference in the shape of your pharynx between the two conditions recommended in this exercise, and that's fine. First, take a large, yawn-like breath and sing a sustained pitch. Next, exhale almost all of that air and sing the same pitch. The quality of phonation should be noticeably warm versus bright and pressed.
2. The next behavior in this discussion is that the larynx does not move straight up and down in the throat. The *thyroid cartilage*, the structure that houses the vocal folds, connects to the topmost cartilaginous ring of the trachea, the *cricoid cartilage*. The cricoid cartilage pulls the larynx along with it as the trachea lowers in the throat. However, the cricoid and thyroid cartilages hinge in the back, and the thyroid cartilage is situated in the frontmost part of the throat. If you touch your thyroid prominence (Adam's apple) and imagine a plumb line dropping from that point, it drops down into the empty space in front of your neck. This means that as the trachea pulls down, the cricoid cartilage follows a somewhat diagonal

path down. Because it hinges with the thyroid cartilage in the back, the front of the thyroid cartilage is free to remain more level with the ground as the trachea pulls it down. Thus, the lowering of the larynx both un-presses the vocal folds at baseline and also opens the anterior space between the thyroid and cricoid cartilages. As the cricothyroid muscles contract to stretch out the vocal folds, they do so by pulling the front of these two cartilages back together. So the additional space between the thyroid and cricoid cartilages appears to increase the range of motion related to that contraction of the cricothyroid muscles.

You can experience this. Take a deep “yawn” breath and glide from your middle voice to a higher pitch. Now blow your air out and try the same thing. You may notice that you strain more to sing the higher pitch when the larynx is higher, the vocal folds are more pressed, and the range of motion between the thyroid and cricoid cartilages decreases. You will also likely need to generate more tracheal pressure to drive this phonation. This will become relevant in a moment.

To be sure, I do not advocate that a lowered larynx is *the optimal preset condition*, especially if that term inspires one to pull the larynx down with muscular force and rigidly keep it there. The larynx freely moves up and down in response to a variety of factors including pitch, timbre, and emotional affect, and an ideal position may not register to the singer as being particularly low. However, it

is important to understand the functional impact of the inhalation of various aspects of the laryngeal setup. Again, we want to end the inhalation with the body exactly where it needs to be to start singing.

Different singers exhale for singing in different ways. I am aware that this is the third rail of voice teaching, so I will only share how I think about this. These ideas are incredibly basic and adapt to many different sounds. When I learned this information, it changed the way I related to previous ways of having been taught to breathe. This includes filling in gaps in my education from teachers who did not address breathing at all.

In general terms, exhalation is facilitated by several driving forces. Muscular force is perhaps the most obvious. Contracting the muscles that compress the abdomen and rib cage forces air out of the body. There is also elastic recoil at work. Depending on how elongated the abdominal muscles become—and different bodies will experience different amounts of abdominal expansion—they will seek to return to their baseline. Both of these actions will push air out of the body automatically, with no additional muscular effort required. If the singer inhales in a way that expands the rib cage, that structure will add its own elastic recoil as well. In addition to muscular and elastic recoil forces, gravity plays a role in respiration dynamics.

Imagine two ways of leveraging the elastic recoil of the abdomen and the rib cage. You could certainly just let the elastic recoil meet its goal of returning these structures to their non-expanded states. To avoid raising the tracheal pressure unmanageably high, you may need to keep some muscles of inhalation somewhat engaged to act as a brake. I observe that this strategy may work especially well with younger singers, given how pliable and accommodating the rest of the system is. However, consider a second option. You could increase the elastic recoil of the rib cage prior to phonation by displacing the abdominal contents in and up. If that sounds

aggressive, keep in mind that this is basically the reverse of the process the diaphragm instigates on the inhalation. Note that this is not a call to engage the muscles of the abdomen that fall along your midline (where I understand a six-pack could hypothetically exist). Those muscles will pull down on the front of your rib cage while the diaphragm is pushed back up in a manner akin to a crunch. We do this when we want to lift something heavy, and it triggers reflexive adduction (Valsalva maneuver) of the vocal folds. The motion I am pointing to does not need to be aggressive at all. It may not even feel like much abdominal engagement.

This may seem counterintuitive, but if the initial expansion of the rib cage was due to displacement of the lower ribs by the diaphragm, and perhaps also by the increased volume of air entering the lungs, your rib cage may remain flexible enough to react to the exhalatory abdominal displacement by expanding further. This accomplishes a few helpful things: (1) It generates a strong elastic recoil driving force in the rib cage for regulating tracheal pressure. (2) It decouples engagement of the abdominal muscles from adduction of the vocal folds, which lowers the chance of triggering a reflexive Valsalva maneuver. And finally, (3) it ensures that the opposing muscles within the rib cage that work to compress or expand that structure are not *also* engaged for postural reasons. The large muscles in the abdomen broadly act on the rib cage, and the small muscles of the rib cage nimbly act on the air in the lungs.

If you would like to explore an extreme but instructive version of this inverse relationship between the abdomen and rib cage, you can carry out an *isomaneuver*. Konno and Mead (1967) suggest taking in a volume of air, holding the breath at the glottis, and displacing “as much volume as possible back and forth between the rib cage and abdomen without flexing or extending the spine.”⁵ As the abdomen decreases in circumference, the rib cage freely expands. As the

abdomen expands, the rib cage automatically falls. Think of the abdomen as easily moving in and out. This is a low-exertion activity. This may be challenging if you habitually resist the expansion of the abdomen or if you hold the rib cage rigid. And to be sure, this is a dramatic illustration of this relationship.

The respiration subsystem may then be understood as two related subsystems. The abdomen is responsible for accommodating the initial descent of the diaphragm. Depending on the sound one wants to make, this may be a relatively small or large descent. The abdomen is also responsible for displacing its own expansion to provide upward force on the diaphragm. This allows the rib cage to expand even further without exerting much muscular or postural effort. Meanwhile, it puts the muscles that compress and expand the rib cage right in the middle of their ranges of motion. This is where they can apply the most force for the least effort.⁶ The elastic recoil of the rib cage may then be leveraged for much longer before muscular effort must kick in, and it becomes the responsibility of the muscles of the rib cage to regulate tracheal pressure. This allows you to remain above the resting expiratory lung volume for much longer as you sing. This is the exhalatory threshold, where elastic recoil forces stop and muscular forces must take over.

Many ways of breathing for singing fit this basic scaffold. This is true even if they look a little different in different bodies, or facilitate different sounds based on where the body ends up at the end of the inhalation. Simply expanding the abdomen less before displacing that volume back into the rib cage changes the output.

This exploration of the respiration subsystem ultimately describes a process that *efficiently and maximally expands the rib cage, storing up elastic recoil energy that may then slowly be released*. One could certainly phonate while allowing the air to fall out with no further activity beyond the elastic recoil of the abdomen and ribcage.

This works very well, especially at lower dynamics. One may also phonate while attempting to push the abdomen out. However, biomechanically this will most likely cause the rib cage to fall. Pushing the abdomen out either directly couples abdominal engagement and vocal fold adduction, which risks pressing, or demands that muscular effort manage rib cage expansion. This works, but it does not do so particularly efficiently. If you would like to read more about the relationship between abdominal and rib cage circumferences in respiration, this has been explored in some detail in the research literature. I encourage those who are interested to explore this endnote.⁷

Phonation and Resonation

In the most rudimentary sense, phonation occurs because tracheal pressure overcomes the resistance of adducted vocal folds. Once set into motion, this process self-sustains until either the tracheal pressure falls below a certain threshold or the vocal folds abduct. How much more detail you would like to bring to your understanding of this process likely relates to what you want to do with this information. The scientific literature has zoomed into minute aspects of tissue histology and airflow patterns. This level of detail is likely not immediately necessary to understand the various functional options available to a singer. Just a gentle reminder that as deeply as you think you understand singing, someone has already gone much deeper.⁸

I group phonation and resonation together because I do not believe it is productive to separate them. It is common to imagine that the oscillation of the vocal folds generates an acoustic signal—the “source”—that is subsequently filtered by the vocal tract—the “filter.” I have argued throughout previous chapters that we may do better to consider that the rapid change in pressure of the

supraglottal air mass *per glottal cycle* is the “source” of the voice. This pressure change is the event that instigates or perpetuates the resonant response of the vocal tract. This instigation literally happens in the vocal tract, not in the vocal fold tissue. Instead, the vocal folds act on the air that flows between them and, in doing so, have a profound effect on that supraglottal pressure change. It is in this spirit that I will offer a narrative to describe the process of phonation and resonation, supported by commentary that outlines pedagogically actionable conclusions and perceptual correlates.

Once the process of respiration has compressed the air in the lungs in some manner, the tracheal pressure will rise. With sufficiently adducted vocal folds, this pressure will exceed the phonation threshold pressure, pushing the vocal folds apart. This is true if starting from complete adduction of the vocal folds, but it is also true if the folds are in a more abducted posture likely to generate a breathy sound. This remains true even if attempting to drive oscillation of only the false vocal folds, a choice that has traditionally been most common in growl or distortion techniques.

The displacement of the adducted vocal fold tissue occurs from bottom to top. There may also be an anterior to posterior pattern of displacement, but we will ignore that for now. Depending on how adducted the vocal folds are, the phonation threshold pressure may be high. Depending on registration, more vocal fold tissue may be in the way from bottom to top. As the tracheal pressure starts to push the vocal fold tissue out of the way, that tissue traces a lateral path from the midline of the glottis out to its maximum place of excursion and then back toward the midline. The ability of tracheal pressure to overcome vocal fold adduction is a primary driving force for vocal fold opening in self-sustaining oscillation. I hope this builds some intuition that the nature of vocal fold adduction, the depth of vocal fold contact, and the vocal fold tissue, tone, and composition will all impact the necessary range of tracheal pressure. This means that

pitch, loudness, and timbre all affect and are affected by tracheal pressure.

Once the vocal folds allow air to flow between them (transglottal flow), the pressure regimes below, within, and above the glottis all start to shift. The vocal fold tissue returns from its maximum excursion back to the midline largely because of the release of elastic recoil energy stored in that tissue as they opened. Again, we find a subsystem that reacts to the way in which it receives energy, stores, and then releases that energy according to its own properties. This suggests that the return from the maximum excursion to the midline slows down as that elastic recoil energy is spent. Toward the end of this return path, the pressure between the folds drops, facilitating an acceleration in the contacting speed of the vocal folds, and the contacting wave ripples from bottom to top of the vocal fold tissue. At the moment of contact the tracheal pressure spikes as the supraglottal pressure drops. Depending on the depth of vocal fold contact, the underside (inferior) edge of the folds is both parted first while the superior tissue remains adducted and also starts its return journey back to midline first. This is an additional asymmetry that suggests that the deeper the vocal fold contacting surface, the faster the cessation of transglottal airflow relative to its instigation.

This asymmetrical airflow pattern—a slower rise in flow followed by a faster cessation of flow—helps to shape the flow pattern of the “new” air entering the vocal tract. Recall our earlier exercise stimulating a low pitch on a loud [a]. I suggested that this stimulus will automatically shorten your thyroarytenoids, facilitating a loose mucosal cover and releasing strain on the vocal ligament. The vocal folds will make deeper contact. Inhalation and tracheal pressure will adjust for this sound. The sound will likely gain auditory roughness and some pitch-less noise, and the formants—all faster than the rate of vocal fold oscillation—will exhibit much complexity. That stimulus

then also generates a flow pattern with a slow rise in flow and a much faster cessation of that flow as the thyroarytenoid muscle engagement pushes the inferior portion of each vocal fold to midline and the loose mucosal cover freely moves over the deeper muscle. How high the flow rate rises depends in part on the degree of muscular adduction of the vocal folds. This suggests that a deep breath may—all other things being equal—ultimately allow more air to flow through the glottis every glottal cycle because the vocal folds will un-press a little at baseline. Look back to figure 7.8 in [chapter 7](#) and remember that features of robust flow phonation include higher rates of airflow overall and a faster cessation of that airflow per glottal cycle.

It may seem odd to imagine that this robust sound could be generated by vocal folds that are not adducted as much as possible. Common sense tells us that more effort should produce more power. And in a linear system it likely would. Consider, though, that the kind of vocal fold posture described here allows the loose mucosal cover to slide around as it oscillates. Because the mucosal cover has a wider excursion than more aggressively adducted vocal folds, the elastic recoil of that tissue easily travels to the midline with the bottom-to-top asymmetry, achieving complete closure while transferring more air to the vocal tract and shutting off that airflow more quickly. In this way, shorter vocal folds with a loose cover feature a characteristic rotational motion. The result is a heavier registration with a more robust sound characterized by both warmth (the slower aspects of the airflow) and buzzy brightness (the faster aspects of the subsequent drop in supraglottal pressure).

When the vocal folds are stretched to sing higher pitches, all of these dynamics shift. Remember stimulating our high pitch (below around A4) on a quiet [u] from earlier. I suggested that this stimulus would decrease the engagement of your thyroarytenoid muscles, stretch your vocal folds with your cricothyroid muscles, and hand

over the mucosal wave of self-sustaining vocal fold oscillation from the slack cover riding over the stiff fibers of your thyroarytenoids to the now stiffer vocal ligament. Your vocal fold contacting area will decrease, especially in the bottom-to-top (inferior-to-superior) dimension. This means that the rotational motion lessens as the depth of vocal fold contact decreases, and the tracheal pressure must rise to move stiffer vocal folds. Taken together, this means that the difference between the rate of accelerating the air through the glottis and how fast that airflow drops becomes more sinusoidal—more like mirror images of each other—although it will always be skewed to some extent. This is because the difference in timing between the superior and inferior edges of the vocal folds decreases. The underside of the vocal folds no longer interrupts the process as quickly.

It may seem counterintuitive that this adjustment ultimately generates a sound with less auditory roughness, less pitch-less noise, and with simpler sounding formants. Tissue vibrating under greater tension should make a brighter, louder sound. But it is the effect of the tissue on the air itself that generates the sound. We have to consider that the slower cessation of airflow results in slower rates of change in the supraglottal air mass.

Let us now turn our attention to other aspects of the role of the larynx in phonation. The posterior ends of the vocal folds attach to the movable *arytenoid* cartilages. These cartilages may move in several ways, and they are responsible for both opening the vocal folds while breathing and drawing them together while phonating. They may completely close the posterior end of the vocal folds, facilitating complete closure every glottal cycle. They may partially close, allowing some airflow as the vocal folds contact. These are continuously adjustable structures.

Stimulate a breathy sound and the arytenoid cartilages will move away from midline to partially open the vocal folds. Stimulate a

glottal onset (for example, a /ga/), and they will fully close before you phonate. As with other laryngeal structures, they respond to stimuli. Additionally, the role of the thyroarytenoid muscles in adduction can be significant. When fully engaged, these muscles bulge toward the midline, facilitating complete, deep closure of the vocal folds per glottal cycle. This can generate imbalances if those contracted muscles resist the stretching required to sing higher pitches, rather than gradually releasing as pitch rises. Cracks can occur when those muscles suddenly release, but the sound is typically bright, rough, and loud right up to the crack. As mentioned in the respiration section above, the basic position of the larynx can affect the ability of these structures to move, and, as I will explore now, the shape of the vocal tract can have a strong effect on the flow pulses the vocal folds facilitate.

It is easy to build models based on the parts of a system we can see. Because the larynx is a physical structure, registration models have traditionally been framed around how that structure changes. The tongue and the other articulators are also physical structures, and they affect the sound based on their positions and motions. However, the sound is a result of the behavior of the air itself. The larynx is interesting insomuch as it modulates airflow. The vocal tract and the various articulators are interesting insomuch as they alter the behavior of the air in the vocal tract. The patterns of pressure change in the air are what reach our ears. In a very real way, the voice is the air. Thus, as we explored with phonation above, we need to understand how the vocal tract affects the behavior of the air it contains.

From the vocal tract's point of view, the flow of air into the vocal tract is not welcome. The vocal tract is already full of air that is reluctant to move, especially at the velocities of phonation. New air increases the density—and therefore pressure—of the supraglottal air mass. This might seem paradoxical, but the air used for

phonation does not flow directly out of the mouth. It bunches up when it enters the vocal tract. This creates a plane of higher-pressure air as the rate of transglottal airflow increases. As the flow overcomes the resistance of that air mass, the higher-pressure plane moves forward in the vocal tract. When the vocal folds partially or completely shut off that flow, the space between the vocal folds and the high-pressure region suddenly drops in density. A drop in density is a drop in pressure.

Because the transglottal airflow pattern is asymmetrical, the pressure drops in the supraglottal air mass faster than it rose. Hypothetically, if we were limited to generating sound powered only by the positive pressure generated by the flow, we would make neither bright nor particularly loud sounds. We cannot directly accelerate a change in air pressure fast enough. The response of the air itself, however, is very fast indeed. Displaced air molecules will swing back toward their equilibrium position quickly. Thus, the faster the cessation of airflow, the faster the supraglottal air mass will drop from high to low pressure, because that step involves no muscles or tissue: It is a property of the air itself. As covered in [chapter 3](#), it is this change in the pressure of the supraglottal air mass that excites the rest of the air in the vocal tract. As this change from high to low pressure occurs in a continuous fashion, continuous rates of change in air pressure characterize that curve. We can understand this "source" as a broadband impulse with elements of slow and fast pressure changes. We hear those rates of change as dark and bright. How resistant the air mass is to the flow depends on several factors.

Air in a contained space behaves differently than unbounded air. Air bounded by hard surfaces behaves differently than air bounded by soft, acoustically absorbent surfaces. In simple terms: The smaller the container, the stiffer the air, and the faster its own elastic recoil properties will respond. The harder the walls of the container,

the longer the duration of the resulting oscillation. I have referred to this as “damping” throughout this book. Thus, brightness in the voice appears to relate to making some portion of the vocal tract just small enough—what colleagues of mine and I call *beneficial narrowing*. Ringing and resonant sounds appear to relate to making a part of the vocal tract harder than was softer. Recall the resonance strategies covered in [chapter 10](#). I doubt that Bareilles, Corelli, Pasquale, or Hampson *softened* their tongues or pharyngeal walls to produce those cutting sounds. Or consider approaches to overtone singing as introduced in [chapter 6](#). Practitioners of overtone singing understand that it is the *tone* of the tongue muscles as much as the position in the mouth that helps generate that strong resonance.

Here are two practical ways to think through this: (1) The tongue is generally not useful for singing when completely relaxed. This is an article of faith to some. However, I think it is demonstrable that some approaches to singing benefit from a slightly firm tongue. (2) You can sequence where in the vocal tract the beneficial narrowing falls and, based on the sound, can verify the behavior of potential vocal tract narrowing that you cannot see.

If the narrowing is as far forward as possible at the lips—like a dark [u] or straw phonation—that means the entire air mass from the vocal folds to that constriction point can all pick up a little of the slack from the flow. This is a very warm sound with little to no buzziness or noise, at least relative to more-open sounds. The vocal tract does not strongly fight the transglottal flow, so the entire air mass can act as a long shock absorber. This is so dependably true that you can tease out whether you are unintentionally narrowing further upstream. A buzzy quality on this sound means you are making an additional narrowing further back in the pharynx.

If the narrowing is with the tongue dorsum, you will hear a clear vowel sound; [i], [e], and [æ] demonstrate this well. This sound can have both warmth from the transglottal flow and brightness from the

beneficial narrowing of the tongue against the roof of the mouth. This sound characterizes a neutral “home base” for many popular-styles singers, and it offers a lot of flexibility regarding registration. To identify whether the tongue is the only actor, stick it out of the mouth and sing. If the sound is still bright and buzzy, the narrowing is further upstream in the pharynx. If the sound is dull with the tongue out and transitions to a clear sound as the tongue returns to its position in the mouth, you have found the right amount of effort to leverage this narrowing. If you again think of the vocal tract air mass as a shock absorber, it is now shorter and springs back faster.

It is also possible to narrow more than one part of the pharynx. A humming sound with buzzy brightness, for example, employs pharyngeal narrowing. The nose itself generates a dopey and dull sound as a resonator rather than a bright and edgy sound. This kind of pharyngeal narrowing is common in character voices and in some musical theater aesthetics. Think Kristin Chenoweth. Kerrie Obert associates this pharyngeal narrowing function (irrespective of nasal airflow) with lateral to medial (sides to the middle) narrowing, and it may encourage vocal fold adduction.⁹ It is characterized by its lack of warmth and its abundance of buzziness and noise. This suggests that a smaller amount of air passes through the glottis every cycle, but the cessation of airflow is rapid.

It is also possible to decrease the aperture of the supraglottal space to stiffen the air mass that the transglottal flow encounters. This tends to generate a cutting, ringing sound with both warmth and buzzy brightness. Previously it was thought that a sphincter muscle narrowed the aryepiglottic fold. More recent work by Obert points to the role of the vallecula (a small empty space in the pharynx anterior to the epiglottis) as an antiresonance that can be mitigated by the action of the tongue root.¹⁰ She labels this *anterior-to-posterior narrowing* (front-to-back). If you can slowly move your

entire tongue mass straight back into your pharynx—not down—to close this space while phonating, you may notice a moment when a very clear sound pops into existence. This happens just before the sound turns into a character voice; [æ] is a great vowel to explore this, although it can be mixed into basically anything but a breathy sound. The skillful implementation of this technique requires only that the root of the tongue move back, freeing the rest of the tongue for articulation. This can be explored through the “weird R” exercise suggested by Chadley Ballantyne or the “wawa pedal sound” suggested by Chris Johnson.¹¹ Both of these exercises encourage a little firmness to the tongue and allow you to play with how much “cut” you want in your sound. Crucially, this “cut” does not automatically come at the expense of any pure warmth and does not require aggressive adduction of the vocal folds or a higher laryngeal position. You can imagine that you do not need to restrict your transglottal flow for this function, and that the drop in supraglottal pressure is rapid, strong, and likely drops from a relatively high-pressure value.

Acoustical Interactions

In a speech-pitch range, where the vocal tract dampens significantly per glottal cycle, the interaction of the vocal folds and vocal tract to form the source impulse is typically characterized by the resistance of the vocal tract to the flow itself. This does not cease to be true as pitch and/or intensity rise, but acoustical interactions do play an increasingly important role. Because the resonances of the vocal tract respond to the source impulse per period, because these resonances bounce forward and backward within the vocal tract as they oscillate, and because humans frequently sing in pitch and intensity ranges where some portion of the vocal tract excitation pattern will still be in motion when the next glottal impulse comes,

the pressure (high or low) of a strong resonance may return to the glottis in a constructive or destructive manner. A constructive alignment happens when a low-pressure moment in the oscillation of the resonance meets the closing glottis in the supraglottal air mass. This means that the drop in pressure facilitated by the rapidly contacting vocal folds joins forces with the drop in pressure from the resonance. This alignment generates a stronger supraglottal impulse than either would have alone, bringing power to that resonance and heightening the tone color contribution of the resulting formant. In some cases, the drop in pressure from the resonance may affect the vibratory pattern of the folds themselves.¹²

The basic condition for this alignment to occur is that the resonance progresses through an integer multiple of its oscillating cycle every pitch period. We have traditionally conceptualized these as *harmonic* multiples of the fundamental. You may also conceptualize the fundamental as a *subharmonic* of the resonance, since the vocal folds oscillate once for every two, three, four, etc., oscillations of the resonance.

Mistiming these events does not prevent phonation. We are not brass instruments, and our resonator does not force the oscillation of our vocal folds to occur at only resonant frequencies of the vocal tract. However, the difference between aligning these acoustical relationships and misaligning them does affect both the power of the source impulse and also whether acoustic power accumulates period to period. Especially in higher pitch ranges while singing louder, a noticeable dip in amplitude can occur when passing between well-aligned and misaligned resonance-pitch relationships. We tend to hear these misalignments as a neutralization of the vowel. These consequential amplitude variations tend to be mitigated by singing lower pitches quietly.

Summary

In summary, to sing, the abdomen expands with or ahead of the rib cage and then asymmetrically decreases its circumference to displace that expansion into the rib cage. That expansion maximizes the elastic recoil energy stored in the rib cage. This elastic recoil asymmetrically raises tracheal pressure sufficient to drive the vocal folds apart. The vocal folds also store elastic recoil energy in their tissues, using it to assist their return to the midline. In returning to their midline, led by the inferior edge of the vocal folds, the transglottal flow pattern is asymmetrically skewed. The supraglottal air mass resists that airflow according to the shape of its container and, in doing so, raises pressure of the supraglottal air. This leverages the elastic recoil of the air itself, which facilitates a subsequently faster drop in supraglottal pressure as the transglottal airflow rate drops and the high-pressure air moves down the vocal tract. That drop from high to low pressure propagates through the air mass in the vocal tract as a pressure wave, interacting with various changes in diameter to set up a cascade of resonant responses. At higher pitches and higher amplitudes, the oscillation of these resonances may time well with subsequent drops in supraglottal pressure as the vocal folds contact. This can preserve the amplitude of the resonances already in motion in the vocal tract and facilitate a stronger initial source impulse, affecting the entire spectrum. This elegant nonlinear system can work exquisitely well to produce high output for minimal effort, and changes in one part should affect the behavior of other parts both up- and downstream.

GOOD RULES OF THUMB TO CONNECT PERCEPTION TO FUNCTION

Three good rules of thumb emerge from the above that help one to quickly and heuristically link perception of the radiated sound to the function that generated it. As with any functional correlate to a desired sound, I suggest that you approach stimulating these functions indirectly.

1. The strength of the warm part of the sound is tied to the volume of airflow through the glottis and the resulting high supraglottal pressure. This is the slowest part of the flow and pressure pattern.
2. The strength of the brightest,uzziest, and noisiest parts of the sound are tied to how quickly the supraglottal air mass drops in pressure. This may be caused by firmer vocal fold adduction; by a stiffer, more reluctant supraglottal air mass; or by a well-timed resonance returning to the glottis as the vocal folds contact. This is the fastest part of the flow and pressure pattern, and it is especially susceptible to narrowing in the pharynx.
3. The clear and potentially cutting part of a vowel that is both pure and resolved is related to the tone and position of the tongue. This action does not describe a use of the tongue that would ever

be characterized as aggressive, tense, or any variety of *pushing the tongue down*. To bring more of this sound into the voice, lightly firm the tongue slightly, slowly move the tongue mass directly back in the mouth, or ultimately use the root of the tongue to *gently* close the vallecula.¹³ This is a common gesture in higher energy singing, and it likely happens despite a sense that the throat is otherwise “open” or “unconstricted.”

NOTES

1. Cornelius Reid, "Functional Vocal Training," in *Journal of Orgonomy* 4, no. 2 (1970): 232.
2. Cornelius Reid, "Sixty Years on the Bench," in *The Modern Singing Master: Essays in Honor of Cornelius L. Reid*, accessed 20 September 2024, <https://corneliusreid.com/wp-content/uploads/2014/08/60-years-on-the-bench.pdf>, 11.
3. Ingo R. Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory," in *Journal of the Acoustical Society of America* 123, no. 5 (May 2008).
4. Johan Sundberg, "Breathing Behavior While Singing," in *Journal of Singing* 49, no. 3, (January/February 1993): 4.
5. K. Konno and J. Mead, "Measurement of the Separate Volume Changes of Rib Cage and Abdomen during Breathing," in *Journal of Applied Physiology* 22, no. 3 (March 1, 1967): 407–22, <https://doi.org/10.1152/jappl.1967.22.3.407>.
6. Ingo R. Titze and Katherine Verdolini Abbott, *Vocology: The Science and Practice of Voice Habilitation* (Salt Lake City: National Center for Voice and Speech, 2012), 22–24.
7. Sally Collyer, "Breathing in Classical Singing: Linking Science and Teaching" (paper presented at the International Symposium on Performance Science, Melbourne, Australia, 2009), in *Proceedings of the International Symposium on Performance Science 2009*, 153–58 (Utrecht: European Association of Conservatoires [AEC], 2009); Sally Collyer, Dianna T. Kenny, and Michaele Archer, "The Effect of Abdominal Kinematic Directives on Respiratory Behaviour in Female Classical Singing," in *Logopedics Phoniatrics Vocology* 34, no. 3 (January 2009): 100–110, <https://doi.org/10.1080/14015430903008780>; S. D. Foulds-Elliott, William Thorpe, S. J. Cala, and Pamela Jane Davis, "Respiratory Function in Operatic Singing: Effects of Emotional Connection," in *Logopedics Phoniatrics Vocology* 25, no. 4 (January 2000): 151–68, <https://doi.org/10.1080/140154300750067539>; Mattias Heldner, Marcin Włodarczak, Peter Branderud, and Johan Stark, "The RespTrack System" (paper presented at the 27th International Congress on Sound and Vibration, Sønderborg, Denmark, 2019), 16–18, <https://www.diva-portal.org/smash/get/diva2:1467649/FULLTEXT01.pdf>; Thomas J. Hixon, Michael D. Goldman, and Jere Mead, "Kinematics of the Chest Wall during Speech Production: Volume Displacements of the Rib Cage, Abdomen, and Lung," in *Journal of Speech and Hearing Research* 16, no. 1 (March

1973): 78–115, <https://doi.org/10.1044/jshr.1601.78>; Thomas J. Hixon, Jere Mead, and Michael D. Goldman, "Dynamics of the Chest Wall during Speech Production: Function of the Thorax, Rib Cage, Diaphragm, and Abdomen," in *Journal of Speech and Hearing Research* 19, no. 2 (June 1976): 297–356, <https://doi.org/10.1044/jshr.1902.297>; Jeannette D. Hoit, Christie L. Jenks, Peter J. Watson, and Thomas F. Cleveland, "Respiratory Function during Speaking and Singing in Professional Country Singers," in *Journal of Voice* 10, no. 1 (January 1996): 39–49, [https://doi.org/10.1016/S0892-1997\(96\)80017-8](https://doi.org/10.1016/S0892-1997(96)80017-8); K. Konno and J. Mead, "Measurement of the Separate Volume Changes of Rib Cage and Abdomen during Breathing," in *Journal of Applied Physiology* 22, no. 3 (March 1, 1967): 407–22, <https://doi.org/10.1152/jappl.1967.22.3.407>; R. Leanderson and Johan Sundberg. "Breathing for Singing," in *Journal of Voice* 2, no. 1 (January 1988): 2–12, [https://doi.org/10.1016/S0892-1997\(88\)80051-1](https://doi.org/10.1016/S0892-1997(88)80051-1); Sauro Salomoni, Wolbert van den Hoorn, and Paul Hodges, "Breathing and Singing: Objective Characterization of Breathing Patterns in Classical Singers," in ed. Charles R. Larson, *PLOS ONE* 11, no. 5 (May 9, 2016): e0155084, <https://doi.org/10.1371/journal.pone.0155084>; Johan Sundberg, "Breathing Behavior While Singing," in *Journal of Singing* 49, no. 3, (January/February 1993): 4–9, 49–51; Monica Thomasson, "Belly-in or Belly-out? Effects of Inhalatory Behaviour and Lung Volume on Voice Function in Male Opera Singers," in *Department for Speech, Music and Hearing Quarterly Progress and Status Report* 45, no. 1 (2003): 61–74; Peter J. Watson and Thomas J. Hixon, "Respiratory Kinematics in Classical (Opera) Singers," in *Journal of Speech, Language, and Hearing Research* 28, no. 1 (March 1985): 104–22, <https://doi.org/10.1044/jshr.2801.104>; Peter J. Watson, Thomas J. Hixon, Elaine T. Stathopoulos, and Daniel R. Sullivan, "Respiratory Kinematics in Female Classical Singers," in *Journal of Voice* 4, no. 2 (January 1990): 120–28, [https://doi.org/10.1016/S0892-1997\(05\)80136-5](https://doi.org/10.1016/S0892-1997(05)80136-5).

8. Zhaoyan Zhang, "Mechanics of Human Voice Production and Control," in *Journal of the Acoustical Society of America* 140, no. 4 (October 2016): 2614, <https://doi.org/10.1121/1.4964509>.

9. Kerrie Obert, Jihyeon Yun, Donna Erickson, Matthew Reeve, Helen Rowson, and Klaus Møller, "Voice Quality: Interactions Among F0, Vowel Quality, Phonation Mode and Pharyngeal Narrowing," in eds. O. Niebuhr and M. Svensson Lundmark, *Proceedings of the 13th Nordic Prosody Conference: Applied and Multimodal Prosody Research* (Sønderborg, Denmark, 2023): 190–99, <https://doi.org/10.2478/9788366675728-016>.

10. Kerrie Obert et al, "Voice Quality," 194.

11. Please see the online resources for Chapter 11 found at <https://www.embodiedmusiclab.com/hearing-singing> or <https://www.bit.ly/HearingSinging>.
12. Ingo R. Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory," in *Journal of the Acoustical Society of America* 123, no. 5 (May 2008): 2733–49, doi: [10.1121/1.2832337](https://doi.org/10.1121/1.2832337); PMID: 18529191; PMCID: PMC2811547.
13. Kerrie Obert et al., "Voice Quality": 190–99.

12

Pedagogic Practices and Functional Listening Examples

As with [chapter 11](#), this chapter is unlike the preceding material in this book. Instead of providing a narrative exploration through prose, this chapter lists important pedagogic practices and perceptual signifiers of various functional adjustments. I will also provide a list of specific ways I implement the psychoacoustical model that has been thoroughly explored in the preceding chapters.

ADJACENT OR INTERMEDIATE FUNCTIONS AND EXERCISES

One can begin to construct *adjacent* or *intermediate functions* to target the desired function. This describes a specific sound you can stimulate that targets the *missing* timbral aspects of your target function. It will likely sound less like your target function than other options, but it will more quickly get you to your target function. I hesitate to write these down for fear that you will imagine that this is the extent of the applications. This work is not static; it changes and shifts and responds to the endless variables in the voice studio. Here are a few examples that I find myself using fairly regularly. I frequently use “the middle-voice slide” and the first step of “there are no registers, only qualities” as initial diagnostic exercises with most singers.

Airflow Is Warmth / Warmth Is Airflow

If you want to make a robust treble sound but are gripping somewhere to generate enough “cut” in the sound, work a function first that addresses the airflow. It is the airflow during the open phase of each glottal cycle that generates the slower, warm components of the sound. This will also create a high-amplitude oscillation of the fundamental, which will amplify the entire voice. Take a released breath first. Then make a slightly fuzzy (I like the term *gauzy*), medium-intensity, neutral sound. This will abduct the folds enough so they let enough airflow through. They will also then be more likely to allow the rotational motion of the vocal fold cover to occur once you bring more energy to the system. Once that is established, work to ramp up intensity and brightness by bringing a

narrowing into the vocal tract without changing the amount of laryngeal effort.

Going on a Brightness Hunt

I wrote in [chapter 11](#) that brightness is caused by a beneficial narrowing somewhere in the vocal tract. Even if we feel that our throat is open, in artistic singing there is always a maximal narrowing somewhere. In general terms, that can be at the tongue dorsum, in the pharynx, at the base of the tongue, or at the glottis itself. If you are having issues disentangling one possible narrowing with another, stick the tongue out. Removing the tongue as an agent of narrowing tends to immediately shine a light on whether a structure further upstream is unintentionally narrowed. Ideally you can make sounds that use different narrowings at different times.

The Middle-Voice Slide

If you have issues singing in your middle voice and want to blend your strong lower register with your weaker higher register, slide from a loud, buzzy [a] on a low pitch to a quiet, pure [u] on a high note. I like B_b3 to F4, with chromatic iterations down a whole step and up to a third higher. As you change pitch, you must simultaneously move along the loudness and timbre continuums. This exercise requires a gradual decrescendo up to the higher pitch and a gradual crescendo down to the lower pitch. Likewise, explore every vowel from [a] to [u] on the way up and [u] to [a] on the way down. If your voice flips, choose just one control factor to adjust; only change the loudness or the vowel. You will likely be able to move your flipping pitch up and down, depending on the balance of your control factors. Eventually you will be able to eliminate the flip entirely. You will know the right lower pitch function by its warmth,

clarity, and buzziness. You will know the right upper pitch function by its purity, warmth, and lack of buzziness.

The Terrible Treble Ah

If you want to bring fullness into the bottom of the treble staff on an open vowel, make sure you have both sufficient airflow and a beneficial narrowing. In most cases exploring the American-R sound ([ɹ] or [ə̯]) in *bird*, *world*, or *squirrel* or a clear [y] will bring power to your voice in this pitch range. Move the tongue very slowly toward the [a] position. If the sound suddenly becomes dull or weak (loses a clear and pure quality or loses all buzziness), you most likely moved the top of your tongue too far away from the roof of your mouth or relaxed your tongue too much. Start over, and see how small a change is needed to reach your target sound.

Same Shape / New Sound for Non-Trebles

If you want to bring your lower voice into a higher range with balance, practice with ascending slides. For example, if you want to sing [a] on A♭3–C4–E♭4 but spread and strain to the top, try this. Take a released breath and allow the larynx to follow whatever path the breath leads it on. Slide from Ab3 to Eb4. Imagine that you are stretching your folds out by tilting the thyroid and cricoid cartilages to gently close the space between them without raising the sense of tracheal pressure significantly. Let the sound be easy, warm, move from [a] to [o], and slightly decrescendo to the top. Once you have secured this, simply open your mouth from the [o] to the [a] shape without anticipating a dramatic change. You may be surprised how warm and easy a sound you can produce on that pitch with that shape if you first rebalance the registration. Once this is established, practice that slide from one target [a] timbre to the other without

the intermediate [o]. Repeat at higher pitches and dynamics as is useful.

Same Shape / New Sound for Trebles

If you want to ascend the treble staff with a classical sound—or explore registration options as a musical theater or contemporary singer—you may find it helpful to execute the previous exercise in a higher octave. Target a strong [a] on F4 and slide up to [o] on C5. Do not slide up to the [o] with the force of [a]. Aim to directly sing the [o] without excess effort. It is not an [a] restricted by the [o]. Done well, you will notice that the [o] is a simpler sound than the [a], strongly characterized by the pure, warm, |~o| tone color of the fundamental. Once you establish this, slowly open from the [o] to the [a] shapes without changing your registration. You will notice that the [a] shape is capable of producing a full and warm sound even though it is brighter and buzzier on the lower pitch. Ascend in half steps, allowing the intermediate upper vowel shape to open as pitch rises. It will be happier as an [ɔ] toward the top of the treble staff and will open further to an [ɑ] or [a] shape by G5 or A5. Anticipate that these shapes will sound warmer and less buzzy than they do at lower pitches.

Loud and Exciting Is Not Enough

If you can only sing higher notes loud, and your sound is exciting but feels unsustainable or over-pressurized, consider practicing generating brightness with vocal tract shaping rather than higher tracheal pressure and greater muscular adduction of the vocal folds. Singing at a lower volume allows the thyroarytenoid muscles to lengthen slightly, which allows the tissue to stretch and vibrate faster without raising the tracheal pressure into a problematic range. Establish a warm, gauzy, neutral sound first. Then try to make very

easy, clear sounds at that lower volume. The weird R, wawa pedal, tongue-backing, and pharyngeal narrowing strategies introduced in [chapter 11](#) all work. Once you have established easy clarity in the sound, imagine that your goal is simply to ramp up the energy you put into the system. The system will respond and similarly ramp up its nonlinear behaviors as the amount of energy transferred from each subsystem to the next increases and those downstream subsystems *come alive* as they temporarily store and release elastic recoil energy.

There Are No Registers, Only Qualities

You can sing through most of your usable pitch range in several different modes or overall timbral adjustments. I encourage you to:

1. Work a quiet, gauzy neutral [ma] function starting in your lower range and extending over whatever break you think you have. I recommend wide-ranging arpeggios starting on D3 for non-trebles and A3 for trebles.
2. Work a slightly fuller version of this without the gauzy quality.
3. Sing a dull, neutral sound like an [ʌ] or [o].
4. Sing a bright vowel like [e] with the airflow and warmth of the dull sound.
5. Sing a bright [æ] without that airflow and warmth. Explore these qualities over wide ranges.

Your goal is **to not change your basic vocal quality with pitch**. For example, do not start with a gauzy [ma] and slide up an octave

to a bright [æ]. That is for artistic singing. Spend time noticing that it is a choice to do so, not an obligatory coordination. Find out whether you can sing a fast arpeggio on a gauzy [ma]. You may be surprised that the continual stimulus of that sound allows you to sing far higher than normal without strain or a sensation of a register break. Practice exercises 3 to 5 quietly and loudly. Practice them with a crescendo, with a diminuendo, and with equal power as you ascend. Use a lot of low-pitch glides to avoid *hiding* registration transitions. Really imagine that you can sing with the same quality over a wide pitch range. Please check out the work of the Complete Vocal Institute for more ideas,¹ but you can also generate your own inventory of varying voice qualities.

WHAT ARE THE PERCEPTUAL QUALITIES OF VOCAL REGISTERS?

The stereotypical perceptual signifiers of *chest voice* versus *head voice* or *falsetto* can be understood in terms of the presence of buzz (auditory roughness), which is present in the former and typically less present in the latter.

Whoop (hoot) acoustic registration can be understood in terms of the changing tone color of $1f_0$ and the perceptually separable complexity of higher harmonics.

Mix can be understood as the inclusion of the auditory roughness that defines *chest voice* into a laryngeal adjustment that permits a faster $1f_0$ (higher pitch) without undue effort.

True belting may have a strong second harmonic ($2f_0$) in a spectrum, and mix-belting may similarly have a strong third harmonic ($3f_0$), but those harmonics convey a simple, pure, somewhat neutral quality. The defining quality of these sounds is their buzzy brightness higher in the spectrum. That is why they evoke qualities of *chest voice*.

WHAT ARE THE PERCEPTUAL QUALITIES OF REGISTRATION?

All vowels have additional, complementary colors other than their vowel defining color. These may be darker or brighter. Pitch changes the qualitative character—the pitch and auditory roughness percepts—of the characteristic and complementary parts of a vowel. Different vowels tend to be characterized by the tone colors of different parts of the spectrum. The vowels [u] and [o] are generally most strongly defined by their lowest formant (F_1). Most other cardinal vowels are defined by the second spectral peak (F_2).

The shifts that occur with /a/-like vowels on the treble staff (*in close timbre*, according to Bozeman)² can be understood in terms of the effect of merging two separate vowel formants into a single radiated peak. This merging can be understood in the time-and-pressure domain as the result of the duration of timbral evolution prior to the next glottal closing. This does not happen at the same pitch point for brighter, higher second formant vowels like [i] and [e]. One may sing those vowels with a clear vowel identity much higher than the pitch of the first resonance (f_{R1}) so long as one opens the mouth or spreads the lips to track the first resonance higher than the pitch, while keeping the relationship between the tongue and the roof of the mouth the same. It feels somewhat like “re-raising” your tongue, which would otherwise have traveled down as the mouth opens.

Treble voices in particular exhibit a dependable pitch- and vowel-dependent registration trajectory. Below the frequency of the first resonance (f_{R1}), the fundamental ($1f_o$) contributes a warm tone color that is unnecessary for vowel identification. Once you sing the pitch of that resonance, the fundamental ($1f_o$) contributes a

complementary warm vowel color in most cases. As pitch rises, and assuming the singer is opening to speed up that first resonance, eventually what had been two vowel formants plus a warmer color becomes a single formant dominated by $1f_0$. Higher pitches are phenomenologically simpler stimuli and obligatorily generate simpler resulting percepts.

The vowel-like sounds of high-pitched singing (think the C5 of a classical tenor or Eb5 high belt of a singer of contemporary or popular styles) draw on workable options from a limited set of potentially resonant vocal tract shapes at that pitch. The vowel heard (if we are determined to call it that) is a result of how our ears react to the resulting pressure wave. The same vocal tract shape two octaves lower will produce a different sound.

Above the treble staff, classical treble singers basically have two resonant vowel options: the one that sounds like a simple version of [a] and that vowel with added brightness because the tongue is higher or further back.

The singer's formant cluster³ does not trigger the same qualitative percept in a listener as pitch rises. The contrasting quality of an operatic tenor above C5 and at C5 and below can be understood in terms of the auditory roughness of the SFC. It largely dissipates above C5 as those harmonics fall below $5f_0$. Treble voices that are missing the clustered aspect of the singer's formant can nevertheless produce perceptually-relevant brightness.⁴

The “narrowing” around C5 or D5 for a classical treble singer can be explained as the tone color contribution of $1f_0$ coming to power while it contributes a darker tone color than it will above the treble staff.

Overtone singing can be reconsidered in the time-and-pressure domain, as opposed to the current dominant perspective offered by the frequency-and-time domain. A strong resonance at an integer

multiple of the fundamental can be understood as though it is a pure tone riding the carrier wave of $1f_0$ for the duration of each period.

The tone color contribution of $1f_0$ can account for many of the differences between most vowel sounds pitched above the treble staff and those on or below it. Above the staff, it is rare to encounter a weakened $1f_0$ that is both darker than and incidental to the vowel color. Below the staff, $1f_0$ falls below the frequency center of the lowest vocal tract resonance. Genre-appropriate registration can be thought of as the delay of or transition to $1f_0$ coming to prominence as a strong tone color as pitch rises through the treble staff. This is part of why classical treble singing on the staff can be so warm and contemporary singing can be so bright, even though neither generates as complex a percept as found at lower pitches.

THOUGHTS ON HOW PERCEPTUAL QUALITIES POINT TO VOCAL FUNCTION

At a low pitch, the two-timbre model separation of pure and buzzy may be the most nuance you can bring to functional hearing. On the treble staff, you may be able to more easily hear the separate warm fundamental, the pure and clear peak of a vowel-defining formant, and the general buzziness of higher frequency harmonics. When this buzzy high-frequency energy is clustered in a narrow range, it sounds different than when it is spread out. Above the treble staff the tone color of the fundamental will likely dominate the sound, with the potential for more timbrally simple buzzy brightness.

If you teach unamplified singers in a small studio, the warm, buzzy, and possibly bright noisy parts of their voice will have to be a little overrepresented to you to sound balanced in a concert hall at a distance.

If you teach amplified singers, make sure you understand the difference between warmth and brightness acoustically and the way that an audio signal chain reshapes the spectrum. Train for the amplified result, but keep in mind that a technically secure voice still has a stable unamplified sound.

One of the most useful voice functions is to sing in a chest voice pitch range but easily and quietly enough that you do not generate auditory roughness. You can practice this function across wide pitch ranges. We do not have a good word for this function, but it is the scaffold on which all mixed registration is built.

Almost all classical voice registration utilizes a mix of some sort. The exceptions tend to be either lower pitched sounds or very high and exciting “money notes.” However, even those sounds exist in the context of a voice that is balanced enough to gradually adjust registration as pitch, loudness, and timbre change.

All things being equal, lower sung pitches will be more timbrally complex than higher sung pitch across multiple parameters. For all but the lowest of basses, much of the singing range crosses into the pitch range where speechlike qualities start to evaporate. This might include the complexity of two vowel formants, the presence of a warm |~u| tone color below those vowel formants, the presence of buzziness or noise in brighter parts of the spectrum, or the reliance on brightness to carry the power of the voice.

If singing in a range where the acoustic interactions of the vocal tract start to influence phonation, no resonance typically comes close to the amplitude and power of the first resonance (f_{R1}), specific second resonance (f_{R2}) strategies notwithstanding. Ironically, our ears are typically more sensitive to the bright oscillation of the second and faster resonances (f_{R2} , f_{R3} , etc.). Singing for the brightness first frequently diminishes the warm power of that first resonance. This is more or less appropriate based on aesthetic targets. If singing robust, unamplified music, this may cripple potentially beneficial acoustical interactions.

The buzzy sounds we hear in robust singing are generated by the way we *hear* the higher frequency components of periodic sounds, not the effort we put into making the sound. There will always be a lower frequency spectrum component that has no buzz. It may be quiet, but it is always present. Distorted, growling, and rattling sounds are typically produced by means of a secondary oscillator, either by itself or in combination with the vocal folds. This also means that any vocal fold oscillation can remain both pure and buzzy and fundamentally balanced, even when the entire system is producing a distorted sound.

Every successful unamplified solo singer has some sort of brightness inherent in their voice. This is because the act of producing higher amplitude sounds necessarily creates asymmetries

and skewed energy transfers between subsystems. The sound of the faster aspects of these skewed patterns is brightness. Even classical sopranos singing above the treble staff do not just sound like the tone color of their fundamental unless they are singing extremely quietly. Generally speaking, quieter singing works just fine in an amplified context as long as there is some freely produced brightness in the sound.

Classical trebles who sing with vibrato may feature strong amplitude pulsations in the brighter parts of the spectrum. It will be easier and ultimately less distracting to attend to the relatively more consistent fundamental when evaluating stability, consistency, and intonation.

Breathy sounds are dominated by the warm components of the spectrum. Any brightness is stochastic and tends not to generate buzziness.

SOME BASIC PRINCIPLES AND RECURRING BEHAVIORS

Here I have organized a few additional thoughts that do not fit into our broader functional listening narrative. Consider how they might come up in multiple contexts, and generalize across your teaching population:

Voice registers only exist when you are singing loud enough. The voice is only broken if you use it in a broken manner. Taking these two ideas together, smooth registration involves playing with loudness (and timbre) to navigate challenging pitch transitions. As you are working to balance the system, the louder you sing, the more unbalanced a challenging range will become. This does not mean you have to sing quietly, just that singing quieter is a useful intermediate step in your training.

You can either dump a lot of power into your system or just a little. Both ways sound great, and neither is right. Neither way owns warmth or brightness. You can stimulate a sound that has almost any combination of warmth, brightness, purity, power, buzziness, or noise. When you use more energy, the nonlinear interactions become more powerful. You can depower them by just singing more quietly.

Never get with push what you can first get with stretch. This especially applies to high pitch ranges, which benefit from longer vocal folds. Stretch is easier if the larynx is not high in the throat.

The best way to guide a singer to dynamically rebalance their registration as they sing is to make sure they care about what they are singing. If you seek to communicate, you will bring nuanced cadence and prosody to the text. By definition, we achieve this by dynamically adjusting loudness and timbre. This justifies why we might practice the text separately from the music and practice the

text on a sustained pitch or with speechlike pitch contours before marrying it to the melody.

Language demands ongoing spectro-temporal fluctuations between phonemes. This means that continuous vowels of equal power very quickly stop sounding like language. Enjoy the timbre of the transitions as you move your articulators from sound to sound.

Changes in dynamics are changes in timbre. Getting louder means adding brightness and possibly buzziness, even if that means you must slightly open a close or rounded vowel.

If you open your vowel on the way up, do not forget to close it again on the way down. There was probably an airflow resistance reason you started more close.

Watching just the mouth position of an elite singer may not help you imitate them. An unseen beneficial narrowing further back in the vocal tract may be primarily responsible for the sound you attribute to their mouth shape.

Be careful not to mistake a *behavior* of the voice for the *potential* of the voice. The perceptual signifiers tied to a specific genre almost certainly call on only a subset of potential vocal functions.

There is no technical work outside an understanding of the expressive sound target. We have no idea what we need to work on technically until we know what we need to communicate.

The bottom half of the treble staff is the pitch range of maximal timbral flexibility. The way you choose to register this range is a strong indicator of genre. Assuming you have the physical capacity to sing this high, you can sing in this range in many ways and not hurt yourself.

Some of the loudest sounds you can make require almost no effort, so long as you exploit the asymmetries of energy transfer in the singing body. Singing is a whole-body act, which means no one part of the body has to do all the work.

Very few sounds that most singers will make require you to first press your vocal folds together uncomfortably. And the sounds that do involve muscular medialization of the vocal folds still tend to balance the tracheal and supraglottal pressures so that the resulting phonation is efficient. Play with the liminal space between how much and how little vocal fold adduction is required for the sound you want to make.

If there is a breathy-to-pressed continuum of vocal fold adduction, robust singing is in the middle.

A chaotically noisy exercise—for example, a lip bubble or a voiced fricative—might obscure your ability to hear compromises in vocal function.

The study of voice acoustics does not just ask us to understand how to sing high notes loud. It is a way of understanding the interactive nature of the entire singing system, and it should be a tool to explain any sound the voice can make. If your concept of voice acoustics can only accommodate sounds within one genre, you lack a general understanding.

Incredibly intense sounds can be made with what still feels like balanced laryngeal effort. Adjacent functions will also have balanced laryngeal effort. It is better to ramp the entire system up or down as a whole than to come off your center.

Always understand the functional purpose of an exercise before you give it. Be clear in your mind whether you are stimulating high or low airflow, high or low airflow resistance, deep or shallow vocal fold contact, arytenoid cartilage closure or not. Always know if you are stimulating consequential vocal tract shaping or neutralizing the effect of the vocal tract. You do not need to do more than clarify the resulting pitch, loudness, and timbre for your singer, but you should aim to understand the desired response both declaratively and procedurally.

If you are a classical singer, the thought that you have two defined registers is probably making your chest voice heavy to the top and your headvoice weak at the bottom. Learning to mix and belt helps everything about classical singing.

There are very few functions that cannot get louder or quieter or change pitch up or down. Never ascribe inherent limitations to a vocal function.

NOTES

1. Complete Vocal Institute. *Complete Vocal Technique*. Accessed December 12, 2024. <https://completevocalinstitute.com/complete-vocal-technique/>.
2. Kenneth Bozeman, *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013), 23.
3. Johan Sundberg, "Level and Center Frequency of the Singer's Formant," in *Journal of Voice* 15, no. 2 (2001): 176–86.
4. Ian Howell, "In Search of the Soprano Singer's Formant," presented at the Society for Music Perception and Cognition Conference, San Diego, 2017, YouTube, accessed 8 June 2021, <https://www.youtube.com/watch?v=uCguuEZTi5s>.

Epilogue

How ever you have navigated this book, I am glad to see you once again here at the end. I started working toward the material that ultimately forms this book almost eleven years ago. It has shaped curricula I have designed for both graduate and undergraduate students at multiple conservatories, and it is a significant factor in my applied teaching practice. I have had the chance to watch it unfold in the minds of my own students through exploration, research, and ultimately experience teaching within this framework in the heuristic environment of their own voice lessons. I hope that you now hear singers differently than when you began. I also hope that you can better imagine how to integrate a functional understanding of voice production into your own teaching practice.

You may be inspired to read more about psychoacoustics, especially from authors less well represented in the current voice pedagogy and vocology literature. If you do so, I encourage you to seek connections to the way you listen to voices as you teach people to sing. Much work remains to be done to deepen our community's understanding of functional listening, and if you end up exploring practitioner-led research in this area, more the better.

However, from the start my goal was to offer you tools to better understand and make actionable the sounds our singers make; to help you hear with clarity and understand how to utilize sounds as

functional prompts. There is an entire sensorial world to explore in even a simple sustained vowel. My only hope is that now, at the end of this book, you feel like the answer to the question “Can we talk about the sound of a voice?” is a resounding “Yes.”

Ian Howell
Ann Arbor, Michigan, USA
19 September 2024

Glossary

acoustics: The study of the properties of sound.

airflow: The movement of air. In voice science we differentiate between transglottal airflow (direct) and acoustic flow in the vocal tract (alternating).

aperiodic: A non-repeating pattern.

arytenoid cartilages: A pair of small laryngeal cartilages attached to the back of the cricoid cartilage. The vocal folds attach to these cartilages, and their mobilization moves the vocal folds.

auditory roughness: A buzzy percept caused by the population of critical bands of hearing by multiple stimuli.

autodissonance: See auditory roughness

brightness: The perceptual correlate to frequency.

closed quotient: See contact quotient.

contact quotient (CQ): The percentage of each glottal cycle where the vocal folds remain in contact, thus reducing or eliminating transglottal airflow.

cricothyroid muscles (CT): Laryngeal muscles that tilt the anterior thyroid and cricoid cartilages toward one another. Their effect is to lengthen the vocal folds.

critical bands of hearing: The frequency resolution limitation of the basilar membrane (inner ear).

declarative (knowledge or language): The words we use to describe a motor task in the body. "Know what" language.

usefulness: An emergent property of stable singing whereby the listener is not concerned that the singer is concerned about their

own singing.

equivalent rectangular bandwidth (ERB): A mathematical formula to estimate the critical bands of hearing.

formant: A spectral peak generated by the repeated oscillation of a vocal tract resonance.

formant scaling: The idea that the complete set of formants determines vowel identity by their relationships, rather than the frequency regions they fall into.

formant tracking: Modifying the vocal tract such that the mathematical relationship between vocal tract resonance and vocal fold oscillation remains the same as pitch changes.

Fourier transform: A mathematical process that—by means of time averaging—decomposes a duration of a complex waveform into smaller components, each with frequency, amplitude, and phase. Modern spectrographs use either a discrete Fourier transform (DFT), which acts directly on digital audio samples, or a highly efficient version of the DFT called a fast Fourier transform (FFT).

frequency: The number of repetitions of some pattern per second.

glottis: The space between the vocal folds.

heuristic: A sufficient solution arrived at through experience.

high frequency: Rapid changes (in pressure).

low frequency: Slow changes (in pressure).

missing fundamental: A perceptual phenomenon whereby one perceives the pitch of a periodic tone in the absence of a harmonic component at $1f_0$.

oscillation: Movement back-and-forth around a midline. For example, a child on a swing or pressure waves in the vocal tract.

percept: The cognitive experience of a sensory stimulus.

period: The duration of a pattern that is repeated.

periodic: The regular repetition of a pattern.

procedural: (knowledge or language): The internal knowledge of how to execute a motor task in the body. “Know how” language.

psychoacoustics: The scientific field concerned with sound perception.

pure tone: A reasonably sinusoidal oscillation of pressure that is perceptually similar to a sine wave.

register: A range of pitches in the human voice that are produced by a similar mechanism. Some sources point to a consistent timbral quality across this pitch range as well, but this depends on the genre.

registration: The coordination of various muscular, aerodynamic, and acoustics actors in a voice that may rebalance to find optimal combinations of pitch, loudness, and timbre for a given aesthetic target.

resonance: A term that may be defined in more than one way depending on context. Here it typically means the way a mass (in voice we talk about the vocal tract air mass) continues to oscillate once excited. May also characterize the inherent potential of that system to oscillate if excited or the synchronicity of that oscillation with an external oscillating body.

roughness: See *auditory roughness*.

self-dissonance: See *auditory roughness*.

sine wave: A mathematically pure tone with no harmonics.

sound: A change in pressure that propagates through a medium and undergoes auditory transduction in a living creature.

subharmonic: An extended vocal technique by which asymmetrical oscillations of the vocal folds produce a very low pitch.

thyroarytenoid muscles (TA): The laryngeal muscles that make up the body of the vocal folds. Their effect is to shorten the vocal folds and bulge their inferior edge toward the midline.

tone color: The perceptual experience of brightness.

transglottal airflow: Describing air that flows through the glottis, typically from below.

vocal folds: Paired, multi-layered folds of muscle and other tissue housed within the thyroid cartilage. They may be interposed in the flow of air and set into multiple transient or oscillating patterns.

Appendix A

How to Read Graphs Related to Voice Production

Throughout this book, and indeed much of the voice pedagogy and vocology literature, you will encounter graphs. An author may include them to illustrate an important idea captured in the prose, or they may serve as a reference that you will extract different information from on each visit. Graphs are incredibly compact ways to store information and can either clarify or obfuscate depending on your ability to read them. Or, put in actionable terms, your ability to read them affects your ability to extract important information. This appendix will briefly explore the three most common types of graphs that appear in this book: the waveform, the spectrum, and the spectrogram.

THE WAVEFORM

A waveform (figure A.1) is a *time-series* graph. This means it shows the change of some value over time. If you follow the stock market, you have encountered a waveform in the way the recent performance of the S&P 500 is displayed. In voice science the waveform is used to represent the change of several values over time. The most common is the pressure value of a sound wave, although waveforms are also used to show transglottal airflow patterns, lung volume and oral pressure, electroglottographic signals, respiration band signals, etc. I will only explore audio waveforms here.

The process of converting continuous pressure changes in the natural world into a form that a digital computer can work with requires that we take discrete measurements at a regular time interval. This interval is called the sampling period, and it is usually expressed in its reciprocal: *sampling frequency*. Literally, how many measurements are captured per second. Audio waveforms in voice science are typically sampled forty-eight thousand times per second (48 kHz) or more. An audio waveform is then a continuous series of these measurements where the horizontal axis represents time in seconds (s) or milliseconds (ms) and the vertical axis represents pressure in pascals (Pa). Pressure is always reported in reference to the pressure of the atmosphere, which is labeled P_{atm} . Vertical values above P_{atm} represent moments of compression in the air surrounding the microphone. Those below represent moments of rarefaction (expansion). A standard audio waveform's amplitude is scaled so that it ranges from -1 Pa to $+1\text{ Pa}$, with P_{atm} at the middle zero line. Zoom in far enough and the continuous curve of a digital waveform will turn into a series of individual measurements.

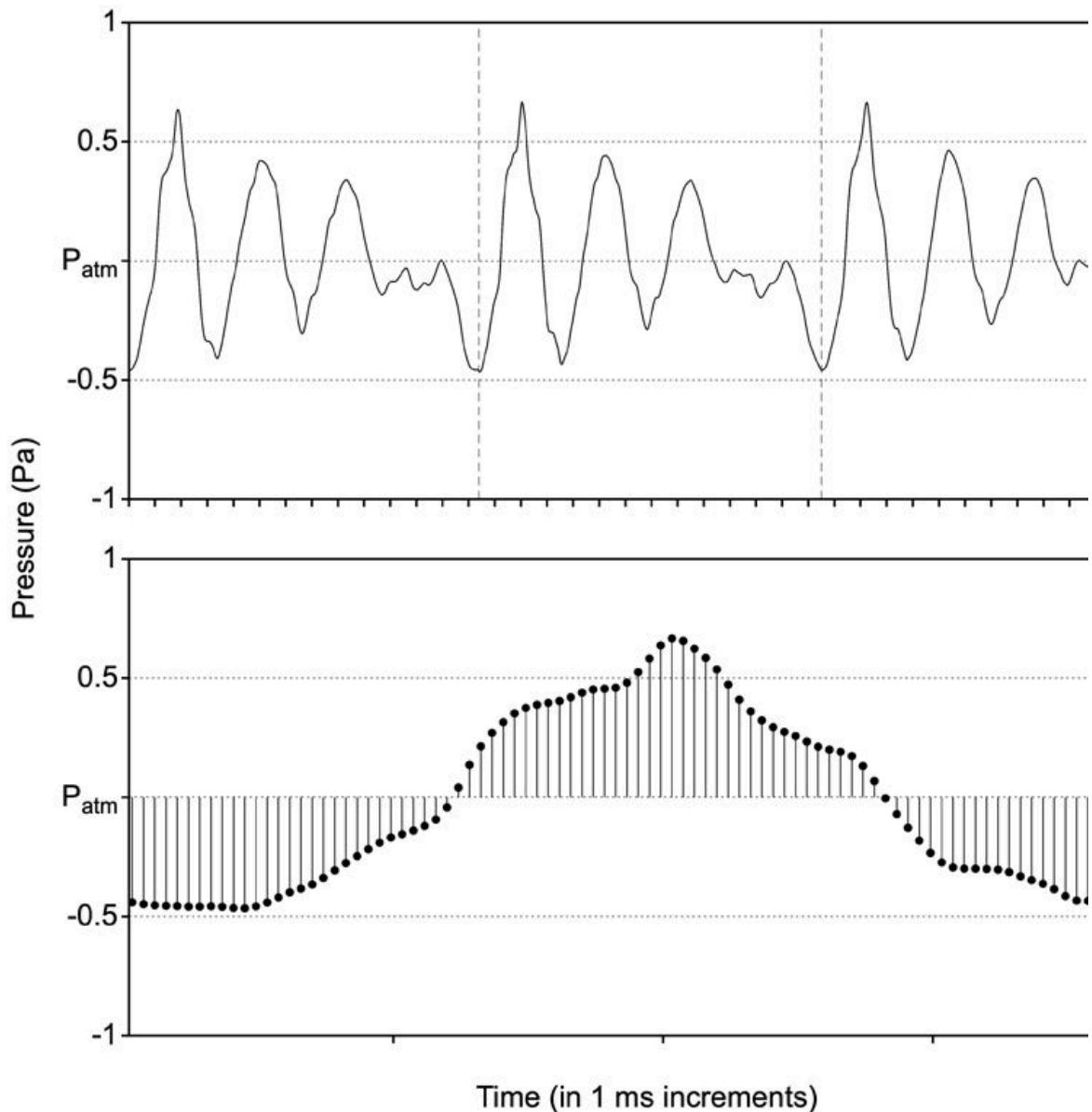


Figure A.1 An audio waveform (above) and a detail showing individual digital samples (below). Recorded by author under controlled conditions.

THE SPECTRUM

A spectrum (figure A.2) is a graph that represents the simple components (simple waveforms with specific frequencies, amplitudes, and phases) that when combined would redraw some duration of a complex waveform. A spectrum is typically the output of a *discrete* or *fast* Fourier transform. The horizontal axis is typically frequency (repetitions per second) in hertz (Hz) and the vertical axis is typically intensity (dB), although you will find these reversed in some programs. You can think of the intensity of each harmonic peak in a spectrum as corresponding to how well a simple waveform of the same frequency lines up with your complex audio waveform. It also takes into consideration the *phase* of that simple waveform, which here points to where it is in its oscillation at the start.

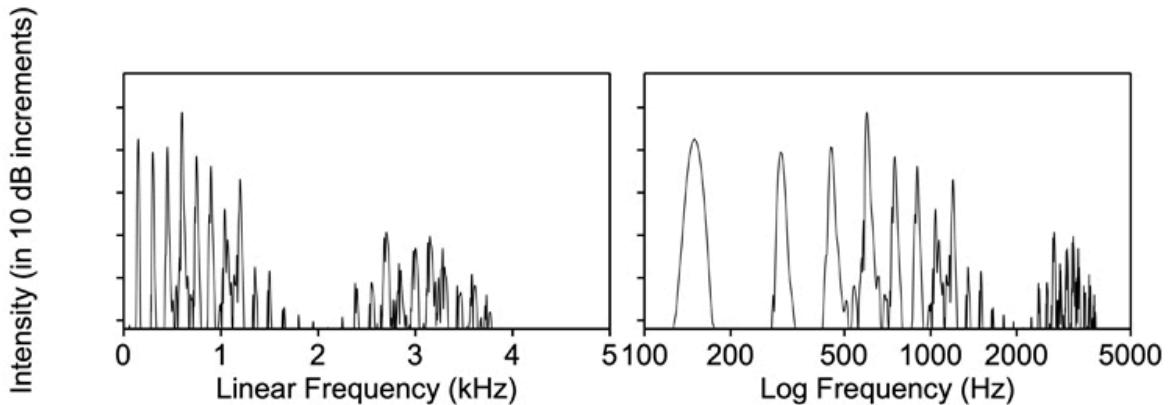


Figure A.2 A narrow-band spectrum generated from a long window of the audio waveform in Figure A.1 Both linear (left) and logarithmic (right) frequency scales are shown.

The frequency axis can be displayed on a linear (figure A.2, left) or logarithmic (figure A.2, right) scale. Linear scales show equally spaced frequencies, while logarithmic scales show equally spaced musical intervals. A change in 100 Hz in a linear scale will keep the same physical spacing throughout the spectrum. In a logarithmic scale, that physical spacing in hertz will get smaller at higher frequencies. In a logarithmic scale a change of an octave (for example) will always retain the same physical spacing regardless of where it is in the spectrum. I use both

linear and logarithmic scales in this book. You will know the frequency scale is logarithmic when there is a piano keyboard in the image.

The Fourier transform may be applied to any duration of an audio file to create a spectrum. When the duration (called a *window*) contains several periods of voicing, the resulting spectrum tends to represent all the pressure fluctuations in the waveform in terms of harmonics of the fundamental. This is true even if the actual pressure pattern within the period does not align well with the pitch period itself—meaning it is not an integer multiple of that fundamental. When the window contains only one or two periods of voicing, the resulting spectrum typically shows no harmonics. The long window spectrum is called *narrow-band*; the short window spectrum is called *wide-band*.

In voice pedagogy and vocology, we tend to measure very long windows of time relative to the duration of a pitch period. VoceVista Video Pro defaults to 83 ms, which will represent almost all periodic voicing with harmonics. In speech and communications and linguistics, both narrow- and wide-band spectra are used. Praat defaults to a very short 5 ms window, which tends to display the supraglottal impulses and the formant frequencies instead of harmonics. All the spectra in this book are narrow-band.

Note that time is not an axis in a spectrum. The entire spectrum represents the entire duration of the window. This means that we do not know *when* in that window the pressure changes represented by a given harmonic occurred. In many cases a harmonic does not point to a continuous pressure change that occurs throughout the entire window. A short window (perhaps one or two pitch periods) will instead show vertical pulses of energy and no harmonics. You can say with greater confidence *when* in the waveform pressure patterns of a given frequency exist, but you lose confidence in the precise frequency.

THE SPECTROGRAM

It is a personal point of irony that the spectrogram is often the first exposure people have to digital audio signal processing, while it is the most complex of the three graphs introduced in this appendix. In concise terms, a spectrogram returns an element of time to the spectrum. A spectrogram is a *time-series* of overlapping spectra. It shows the frequency and intensity content of a series of spectra over time. Whatever time and frequency trade-offs exist with respect to the window duration of the spectra is baked into how a spectrogram represents an audio waveform. This means you may have a wide-band spectrogram (figure A.3, left) or a narrow-band spectrogram (figure A.3, right). It is worth noting that the wide-band view allows you to see the individual glottal impulses as vertical stripes.

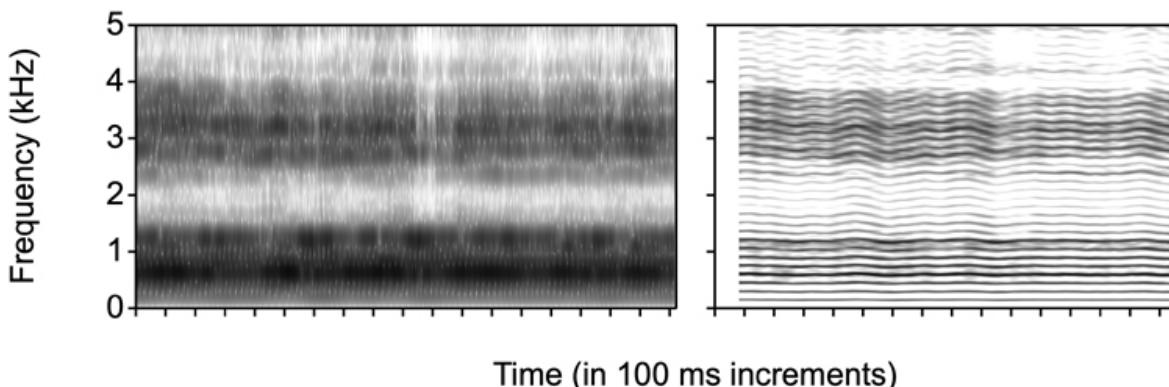


Figure A.3 **Wide-band (left) and narrow-band (right) spectrograms of the audio waveform in Figure A.1.**

You might notice that a narrow-band spectrum looks a little like a mountain range as seen from the side. The spectrogram is then a reorientation so that we are looking down at a time-series of those mountain peaks from above. The horizontal axis now represents time, the vertical axis represents frequency, and a third depth scale (usually shown with a color or gray scale) represents intensity. In the voice pedagogy and vocology literature, we tend to view long-window spectrograms and blend

the information present in the series of spectra to create pleasing images with continuous harmonics.

Appendix B

List of Labs and Website URL

I have created twenty-seven labs for you or your students to work through. Please go to <https://www.embodiedmusiclab.com/hearing-singing> and <https://www.bit.ly/HearingSinging> to sign up for access.

Table B.1 A list of the experiential lab assignments referenced throughout this book.

Table B.1 (Continued)

Lab #	Chapter #	Name
1	2	The Analog Telephone Bandwidth
2	4	Comparing the Tone Color of Different Fundamentals
3	4	Comparing the Timbre of Different Parts of a Spectrum
4	4	Considering the Timbre of Different Onsets, Sustains, and Releases
5	5	Pitch: Hearing Resolved versus Unresolved Harmonics
6	5	Pitch: The Missing Fundamental
7	6	Contrasting the Waveforms of Breathy Singing and Overtone Singing
8	6	Synthesize a Periodic Voice from a Single Vocal Fry Pop

- | | | |
|----|----|--|
| 9 | 7 | Exploring Beats and Auditory Roughness with Two Pure Tones |
| 10 | 7 | Pure and Buzzy Portions of a Complex Spectrum |
| 11 | 7 | Equivalent Rectangular Bandwidth Examples |
| 12 | 8 | Auditory Roughness for Pitches above the Treble Staff |
| 13 | 8 | Vennard's Supposition: The [u] and [o] in the [e] and [i] |
| 14 | 8 | The Shared Tone Color of the Fundamental of [u] and [i] |
| 15 | 8 | Speech Intelligibility of Filtered Spectra |
| 16 | 9 | Intensity-Related Variation in the Pitch Percept of Pure Tones |
| 17 | 9 | Exploring ASTC with Pure Tones and Clusters of Harmonics |
| 18 | 9 | Removing F_2 from Cardinal Vowels |
| 19 | 9 | Comparing the Tone Color and Loudness of $1f_0$ in [u], [i], and [æ] |
| 19 | 9 | Create an F_1 / F_2 Scatterplot |
| 20 | 10 | Experiencing the Same Tone Color in Four Different Positions in the Harmonic Series |
| 21 | 10 | Pass-filtering a Single Tone Color in a Glide from High to Low |
| 22 | 10 | Changing Tone Color of a Formant while Controlling Auditory Roughness and Pitch Resolution |
| 23 | 10 | Changing Auditory Roughness and Pitch Resolution of a Formant while Controlling Tone Color |
| 24 | 10 | Eliciting the ~o Percept with Three Different Pairs of Pure Tones |
| 25 | 10 | Pass-filtering a Treble Singer's $1f_0$ |
| 26 | 10 | Experiencing F_1 and Auditory Roughness of Three Resonance Strategies |
| 27 | 10 | Exploring the Tone Color of Various Vocal Tract Excitation Strategies |
-

Bibliography

- Aaen, Mathias, Cathrine Sadolin, Anna White, Reza Nouraei, and Julian McGlashan. "Extreme Vocals—A Retrospective Longitudinal Study of Vocal Health in 20 Professional Singers Performing and Teaching Rough Vocal Effects." *Journal of Voice* (article in press, 3 June 2022). DOI: 10.1016/j.jvoice.2022.05.002
- Anderson, Julian. "Spectral music." Grove Music Online (2001). Accessed 4 August 2021.
- Anikin, Andrey, Katarzyna Pisanski, and David Reby. "Static and Dynamic Formant Scaling Conveys Body Size and Aggression." *Royal Society Open Science* 9: 211496.
<https://doi.org/10.1098/rsos.211496>
- ANSI. *Psychoacoustic Terminology: Timbre* (New York: American National Standards Institute, 1994).
- Ashcraft, Mark H. *Cognition* (Upper Saddle River, NJ: Pearson Prentice Hall, 2006).
- Austin, Stephen F. "Canaries in the Coal Mine: The Pure Vowel." *Journal of Singing* 68, no. 1 (September/October 2011): 83–85.
- . "Carlo Bassini's 'The Art of Singing, Part 1.'" *Journal of Singing* 66, no. 5 (May/June 2010).
- . "Provenance: The voce chiusa." *Journal of Singing* 61, no. 4 (March/April 2025).
- Bareilles, Sarah. "Goodbye Yellow Brick Road." Recorded May 2013. Track from *Brave Enough: Live at the Variety Playhouse*. Epic Records, 2013.

- Bloom, Harold. *The Anxiety of Influence*, 2nd ed. (New York: Oxford University Press, 1997).
- Borch, Daniel Zanger, Johan Sundberg, P. A. Lindestad, and M. Thalén. "Vocal Fold Vibration and Voice Source Aperiodicity in Phonatorily Distorted Singing." *Department for Speech, Music and Hearing Quarterly Progress and Status Report* (KTH, 2003).
- Bozeman, Kenneth. *Practical Vocal Acoustics* (Hillsdale, NY: Pendragon Press, 2013).
- _____. "Vowel Migration and Modification." *VOICEPrints* 16, no. 2 (November/December 2018).
- _____. *Kinesthetic Voice Pedagogy 2: Motivating Acoustic Efficiency*, expanded edition. (Gahanna, OH: Inside View Press, 2021).
- Bregman, Albert S. *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press, 1994).
- Brown, James. "I Got You (I Feel Good)." From *I Got You (I Feel Good)*. King Records, 1966.
- Brown, Oren. "Resonance and Power." *Discover Your Voice: How to Develop Healthy Voice Habits* (San Diego: Singular, 1996).
- Callaghan, Jean. "Resonance." *Singing and Science: Body, Brain, and Voice* (Oxford: Compton, 2014).
- Chaieb, Leila, Caroline Wilpert, Thomas P. Reber, and Juergen Fell. "Auditory Beat Stimulation and Its Effects on Cognition and Mood States." *Frontiers in Psychiatry* 6 (2015).
<https://doi.org/10.3389/fpsy.2015.00070>
- Chen, C. Julian. *Elements of Human Voice* (Hackensack, NJ: World Scientific, 2017).
- Chiba, Tsutomu, and Masato Kajiyama. *The Vowel: Its Nature and Structure* (Tokyo: Phonetic Society of Japan, 1958).
- Coffin, Berton. *Coffin's Sounds of Singing: Principles and Applications of Vocal Techniques with Chromatic Vowel Chart*, 2nd ed. (Lanham, MD: Scarecrow Press, 2002).
- Cogan, Robert. "Toward a Theory of Timbre: Verbal Timbre and Musical Line in Purcell, Sessions, and Stravinsky." *Perspectives of New Music* 8, no. 1 (Autumn–Winter, 1969).
- _____. *Music Seen, Music Heard: A Picture Book of Musical Design* (Cambridge, UK: Publication Contact International, 1998).

- . *New Images of Musical Sound* (Cambridge, MA: Harvard University Press, 1984).
- Collyer, Sally. "Breathing in Classical Singing: Linking Science and Teaching." Paper presented at the International Symposium on Performance Science, Melbourne, Australia (2009). *Proceedings of the International Symposium on Performance Science 2009* (Utrecht, NL: European Association of Conservatoires [AEC], 2009).
- Collyer, Sally, Dianna T. Kenny, and Michaela Archer. "The Effect of Abdominal Kinematic Directives on Respiratory Behaviour in Female Classical Singing." *Logopedics Phoniatrics Vocology* 34, no. 3 (January 2009). <https://doi.org/10.1080/14015430903008780>
- Complete Vocal Institute. *Complete Vocal Technique*. Accessed December 12, 2024. <https://completevocalinstitute.com/complete-vocal-technique/>.
- Corelli, Franco. "A te, o cara." From Bellini's *I Puritani*. Conducted by Tullio Serafin. EMI Records, 1956. CD.
- Cross, Melissa. "The Zen Of Screaming 2 Trailer!! New DVD!" [3:31]. YouTube. <https://www.youtube.com/watch?v=spZWQwxNKHg>
- Dayme, Meribeth Bunch. "Resonation and Vocal Quality." *Dynamics of the Singing Voice*, 5th ed. (New York: Springer, 2008).
- Denes, Peter B., and Elliot N. Pinson. *The Speech Chain: The Physics and Biology of Spoken Language*, 2nd ed. (New York: W. H. Freeman, 1993; reissued 2015).
- Doscher, Barbara M. *The Functional Unity of the Singing Voice*, 2nd ed. (Lanham, MD: Scarecrow Press, 1994).
- Eilish, Billie. "What Was I Made For?" From *Barbie: The Album*. Darkroom/Interscope Records, 2023.
- Engel, Gustav. "Ueber den Begriff der Klangfarbe." *Philosophische Vorträge* (Berlin). Neue Folge, II. Ser., Heft 12 (1886).
- Engelhardt, V., and E. Gehrcke. "Ueber die Vokal-charaktere einfacher Töne." *Zeitschrift für Psychologie* 115: 91930.
- Fales, Cornelia. "The Paradox of Timbre." *Ethnomusicology* 46, no. 1 (2002).
- Fant, Gunnar. *Acoustic Theory of Speech Production* (The Hague: Mouton De Gruyter, 1970).

- Fluza, Mauro Barro, and Marta Assumpção de Andrada e Silva. "Can Singing with Rasp Be a Healthy Practice?" *Distúrbios da Comunicação* 30, no. 4 (December 2018): 802–808.
<http://dx.doi.org/10.23925/2176-2724.2018v30i4p802-808>.
- Fletcher, Harvey. *Speech and Hearing* (New York: D. Van Nostrand, 1929).
- Foulds-Elliott, S. D., William Thorpe, S. J. Cala, and Pamela Jane Davis. "Respiratory Function in Operatic Singing: Effects of Emotional Connection." *Logopedics Phoniatrics Vocology* 25, no. 4 (January 2000). <https://doi.org/10.1080/140154300750067539>
- Fulop, Sean A. *Speech Spectrum Analysis* (Berlin: Springer, 2011).
- Grassmann, Hermann. "Ueber die physikalische Natur der Sprachlaute." *Annalen der Physik und Chemie* 1 (1877).
- Hampson, Thomas. "Avant de quitter ces lieux." In *Faust* by Charles Gounod. Conducted by Michel Plasson, featuring Cheryl Studer, Richard Leech, José van Dam, and the Orchestre et Chœur du Capitole de Toulouse. Recorded February 15–28, 1991, Halle-aux-Grains, Toulouse. EMI France, 1991. CD.
- Hartmann, William. "Pitch, periodicity, and auditory organization." *Journal of the Acoustical Society of America* 100, no. 6 (December 1996).
- Heldner, Mattias, Marcin Włodarczak, Peter Branderud, and Johan Stark. "The RespTrack System." Paper presented at the 27th International Congress on Sound and Vibration (Sønderborg, Denmark, 2019). <https://www.diva-portal.org/smash/get/diva2:1467649/FULLTEXT01.pdf>
- Heller, Eric J. "Mechanisms of Hearing." *Why You Hear What You Hear* (Princeton, NJ: Princeton University Press, 2013).
- Helmholtz, Hermann L. F. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th ed. (1877), translated by Alexander J. Ellis (New York: Longmans, Green, and Co., 1912).
- Herbst, Christian T. and Jan Švec. "Adjustments of Glottal Configurations in Singing." *Journal of Singing* 70, no. 3 (January/February 2014).

- Herbst, Christian T., Coen P. H. Elemans, Isao T. Tokuda, Vasileios Chatzioannou, and Jan G. Švec. "Dynamic System Coupling in Voice Production." *Journal of Voice*, in press (October 7, 2022).
- Hermann, Ludimar. "Beträge zur Lehre von der Klangwahrehmung." *Pflueger's Archiv* 56 (1894).
- Hixon, Thomas J., Gary Weismer, and Jeannette D. Hoit. "Acoustic Theory of Vowel Production." *Preclinical Speech Science: Anatomy, Physiology, Acoustics, and Perception*, 3rd ed. (San Diego: Plural, 2020).
- Hixon, Thomas J., Michael D. Goldman, and Jere Mead. "Kinematics of the Chest Wall during Speech Production: Volume Displacements of the Rib Cage, Abdomen, and Lung." *Journal of Speech and Hearing Research* 16, no. 1 (March 1973).
<https://doi.org/10.1044/jshr.1601.78>
- Hixon, Thomas J., Jere Mead, and Michael D. Goldman. "Dynamics of the Chest Wall during Speech Production: Function of the Thorax, Rib Cage, Diaphragm, and Abdomen." *Journal of Speech and Hearing Research* 19, no. 2 (June 1976).
<https://doi.org/10.1044/jshr.1902.297>
- Hoit, Jeannette D., Christie L. Jenks, Peter J. Watson, and Thomas F. Cleveland. "Respiratory Function during Speaking and Singing in Professional Country Singers." *Journal of Voice* 10, no. 1 (January 1996). [https://doi.org/10.1016/S0892-1997\(96\)80017-8](https://doi.org/10.1016/S0892-1997(96)80017-8)
- Hopp, Steven L., Michael J. Owren, and Christopher S. Evans, eds. *Animal Acoustic Communication: Sound Analysis and Research Methods* (Berlin: Springer-Verlag, 1998).
- Houston, Whitney. "Whitney Houston—I Will Always Love You (official 4K video)," [4:34]. YouTube.
<https://youtu.be/3JWTaaS7LdU?si=UjKZOM-iZYg28IiB>
- Howard, David M., and Jamie Angus. *Acoustics and Psychoacoustics*, 5th ed. (New York: Routledge, 2017).
- Howell, Ian. "Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color." DMA diss., New England Conservatory of Music, 2016.
- _____. "Necessary Roughness in the Voice Pedagogy Classroom: The Special Psychoacoustics of the Singing Voice." *VOICEPrints*

(May/June 2017).

- _____. "Misreading the Science: Vocal Treatises, Vowels, and a New Framework for Understanding the Female *passaggio*." Unpublished term paper MHST902, New England Conservatory of Music, 2014.
- _____. "In Search of the Soprano Singer's Formant." Presented at the Society for Music Perception and Cognition Conference, San Diego, 2017. YouTube. Accessed 8 June 2021.
<https://www.youtube.com/watch?v=uCguuEZTi5s>
- Humphries, Colin, Einat Liebenthal, and Jeffrey R. Binder. "Tonotopic Organization of Human Auditory Cortex." *NeuroImage* 50, no. 3 (2010). DOI:10.1016/j.neuroimage.2010.01.046
- Köhler, Wolfgang. "Akustische Untersuchungen I." *Beiträge zur Akustik und Musikwissenschaft* 4 (1909). Reprinted in *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 54 (1910).
- _____. "Akustische Untersuchungen II." *Zeitschrift für Psychologie* 58 (1911): 59–140. Reprinted in *Beiträge zur Akustik und Musikwissenschaft* 6 (1911).
- _____. "Akustische Untersuchungen III." *Zeitschrift für Psychologie* 72 (1915).
- Konno, K., and J. Mead. "Measurement of the Separate Volume Changes of Rib Cage and Abdomen during Breathing." *Journal of Applied Physiology* 22, no. 3 (March 1, 1967).
<https://doi.org/10.1152/jappl.1967.22.3.407>
- Lamperti, Giovanni Battista. *The Technics of Bel Canto* (New York: G. Schirmer, 1905), 15. Emphasis added.
- Lass, Norman J., and Charles M. Woodford. *Hearing Science Fundamentals* (St. Louis: Mosby Elsevier, 2007).
- Leanderson, R., and Johan Sundberg. "Breathing for Singing." *Journal of Voice* 2, no. 1 (January 1988).
[https://doi.org/10.1016/S0892-1997\(88\)80051-1](https://doi.org/10.1016/S0892-1997(88)80051-1).
- LeBorgne, Wendy, and Marci Rosenberg. "Resonance and Vocal Acoustics." *The Vocal Athlete*, 2nd ed. (San Diego: Plural Publishing, 2021).
- Lee, Kyoga. "Pitch Perception: Place Theory, Temporal Theory, and Beyond." EE 391 Special Report (Autumn 2004).

https://ccrma.stanford.edu/~kglee/pubs/klee_ee391_fall04.pdf,
accessed 29 June 2021

- Lopez-Poveda, Enrique A., Alan R. Palmer, and Ray Meddis, eds. *Neurophysiological Bases of Auditory Perception* (New York: Springer, 2010).
- Mach, Ernst. "Zur Analyse des Tonempfindungen." *Sitzungsbericht der kaiserlichen Akademie der Wissenschaften in Wien*, Bd. 92, Abt. 2 (1885).
- Mandal, Soumyajit, Serhii M. Zhak, and Rahul Sarpeshkar. "A Bio-Inspired Active Radio-Frequency Silicon Cochlea." *IEEE Journal of Solid-State Circuits* 44, no. 6 (2009).
- McCoy, Scott. *Your Voice: An Inside View*, 3rd ed. (Gahanna, OH: Inside View Press, 2019).
- McKinney, James C. *The Diagnosis and Correction of Vocal Faults* (Long Grove, IL: Waveland Press, 2005).
- McLachlan, Neil M. "Timbre, Pitch, and Music." *Oxford Handbooks*. Published June 2016. DOI:
10.1093/oxfordhb/9780199935345.013.44
- Merman, Ethel. "Everything's Coming Up Roses." From *Gypsy: Original Cast Recording*. Composed by Jule Styne, Stephen Sondheim, and Arthur Laurents. Columbia Masterworks, 1959. LP.
- Miller, Donald G. *Resonance in Singing: Voice Building through Acoustic Feedback* (Princeton, NJ: Inside View Press, 2008).
- _____. "Registers in Singing: Empirical and Systematic Studies in the Theory of the Singing Voice." PhD diss., University of Groningen, 2000.
- Miller, Richard. *Solutions for Singers: Tools for Performers and Teachers* (Oxford, UK: Oxford University Press, 2004).
- _____. *Training Soprano Voices* (New York: Oxford University Press, 2000).
- Mitton, Daniel A. H. "Sung Russian for the Low Male Voice Classical Singer: The Latent Pedagogical Value of Sung Russian." DMA diss., University of Toronto, 2020.
- Monson, Brian, Eric J. Hunter, Andrew J. Lotto, and Brad H. Story. "The Perceptual Significance of High-Frequency Energy in the Human Voice." *Frontiers in Psychology* 16, no. 5 (2014).

- Morton, John, and Alan Carpenter. "Judgement of the Vowel Colour of Natural and Artificial Sounds." *Language and Speech* 5, no. 4 (1962).
- Nair, Garyth. *The Craft of Singing* (San Diego: Plural Publishing, 2007).
- Nearey, Terrance M. "Static, Dynamic, and Relational Properties in Vowel Perception." *Journal of the Acoustical Society of America* 85, no. 5 (May 1989).
- Norman-Haignere, Sam, Nancy Kanwisher, and Josh H. McDermott. "Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex." *The Journal of Neuroscience* 33, no. 50 (December 11, 2013).
- Obert, Kerrie, Jihyeon Yun, Donna Erickson, Matthew Reeve, Helen Rowson, and Klaus Møller. "Voice Quality: Interactions Among F0, Vowel Quality, Phonation Mode and Pharyngeal Narrowing." In O. Niebuhr and M. Svensson Lundmark, eds. *Proceedings of the 13th Nordic Prosody Conference: Applied and Multimodal Prosody Research* (Sønderborg, Denmark, 2023): 190–99.
<https://doi.org/10.2478/9788366675728-016>
- Pasquale, Steven. "Wondering." From *The Bridges of Madison County*: Original Broadway Cast Recording. Composed by Jason Robert Brown. Ghostlight Records, 2014.
- Peterson, Gordon E., and Harold L. Barney. "Control Methods Used in a Study of Vowels." *The Journal of the Acoustical Society of America* 24, no. 2 (March 1952).
- Platt, James, and David M. Howard. "Applied Vocal Acoustics and Acoustic Registration." In Janice M. Chapman and Ron Morris, eds. *Singing and Teaching Singing: A Holistic Approach to Classical Voice*, 4th ed. (San Diego: Plural Publishing, 2021).
- Plomp, Reinier. "Experiments on Tone Perception." PhD diss., Institute for Perception RVO-TNO, 1966.
- . *The Intelligent Ear: On the Nature of Sound Perception* (London: Lawrence Erlbaum, 2002), 23–24.
- Ragan, Kari. "Defining Evidence-Based Voice Pedagogy: A New Framework." *Journal of Singing* 72, no. 2 (2018).

- Reid, Cornelius. "Functional Vocal Training." *Journal of Orgonomy* 4, no. 2 (1970).
- _____. "Functional Vocal Training." *Journal of Orgonomy* 5, no. 1 (1971).
- _____. "Sixty Years on the Bench." *The Modern Singing Master: Essays in Honor of Cornelius L. Reid*. Accessed 20 September 2024. <https://corneliusreid.com/wp-content/uploads/2014/08/60-years-on-the-bench.pdf>
- Roubeau, Bernard, Nathalie Henrich, and Michèle Castellengo. "Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited." *Journal of Voice* 23, no. 4 (2007).
- Rushton, Julian. "Klangfarbenmelodie." Grove Music Online (2001). Accessed 4 August 2021.
- Salomoni, Sauro, Wolbert van den Hoorn, and Paul Hodges. "Breathing and Singing: Objective Characterization of Breathing Patterns in Classical Singers." Charles R Larson, ed. *PLOS ONE* 11, no. 5 (May 9, 2016). <https://doi.org/10.1371/journal.pone.0155084>
- Siedenburg, Kai, and Stephen McAdams. "Four Distinctions for the Auditory 'Wastebasket' of Timbre." *Frontiers in Psychology* 8, article 1747 (2017). <https://doi.org/10.3389/fpsyg.2017.01747>
- Simone, Nina. "Feeling Good (Official Video)." YouTube video, 3:56. Published by Verve Records, July 24, 2013. <https://youtu.be/oHRNrgDIJfo>.
- Slawson, Wayne. *Sound Color* (Berkeley: University of California Press, 1985).
- Story, Brad. "The Vocal Tract in Singing." Graham Welch, David M. Howard, and John Nix, eds. *The Oxford Handbook of Singing* (Oxford, UK: Oxford University Press, 2019).
- Straus, Joseph N. "The 'Anxiety of Influence' in Twentieth-Century Music." *The Journal of Musicology* 9, no. 4 (1991).
- Stumpf, Carl. *Tonpsychologie*, vol. 2 (Leipzig: S. Hirzel Verlag, 1890), 514–43.
- Sundberg, Johan. *The Science of the Singing Voice* (DeKalb: Northern Illinois University Press, 1987).

- _____. "Articulatory Interpretation of the 'Singing Formant.'" *Journal of the Acoustical Society of America* 55, no. 4 (1974).
- _____. "Research on the Singing Voice in Retrospect." *TMH-QPSR, KTH* 45, no. 1 (2003): 011–022.
- _____. "Level and Center Frequency of the Singer's Formant." *Journal of Voice* 15, no. 2 (2001). [https://doi.org/10.1016/S0892-1997\(01\)00019-4](https://doi.org/10.1016/S0892-1997(01)00019-4)
- _____. "Breathing Behavior While Singing." *Journal of Singing* 49, no. 3 (January/February 1993).
- Švec, Jan G., Harm K. Schutte, and Donald G. Miller. "A Subharmonic Vibratory Pattern in Normal Vocal Folds." *Journal of Speech and Hearing Research* 39, no. 1 (1996): 135–43. DOI: 10.1044/jshr.3901.135
- Tallon, Tina. "A Century of 'Shrill': How Bias in Technology Has Hurt Women's Voices." *New Yorker* (September 3, 2019).
<https://www.newyorker.com/culture/cultural-comment/a-century-of-shrill-how-bias-in-technology-has-hurt-womens-voices>
- Ternström, S. O. "Hi-Fi Voice: Observations on the Distribution of Energy in the Singing Voice Spectrum above 5 kHz." *The Journal of the Acoustical Society of America* 123, no. 5 (2008): 3171–76.
- Thomasson, Monica. "Belly-in or Belly-out? Effects of Inhalatory Behaviour and Lung Volume on Voice Function in Male Opera Singers." *Department for Speech, Music and Hearing Quarterly Progress and Status Report* 45, no. 1 (2003).
- Titze, Ingo R. *Principles of Voice Production* (Iowa City: National Center for Voice and Speech, 2000).
- _____. "Why Do Close Harmonies and Dissonances Sound Rougher at Low Pitches than High Pitches?" *Journal of Singing* 73, no. 4 (March/April 2017).
- _____. "Human Speech: A Restricted Use of the Mammalian Larynx." *Journal of Voice* 31, no. 2 (March 2017).
- _____. "Nonlinear Source-Filter Coupling in Phonation: Theory." *Journal of the Acoustical Society of America* 123, no. 5 (May 2008). DOI: 10.1121/1.2832337; PMID: 18529191; PMCID: PMC2811547

- Titze, Ingo R., et al. "Toward a Consensus on Symbolic Notation of Harmonics, Resonances, and Formants in Vocalization." *The Journal of the Acoustical Society* 137, no. 5 (May 2015).
- Titze, Ingo R., and Sung Min Jin. "Is There Evidence of a Second Singer's Formant?" *Journal of Singing* 59, no. 4 (2003).
- Titze, Ingo R., and Katherine Verdolini Abbott. *Vocology: The Science and Practice of Voice Habilitation* (Salt Lake City: National Center for Voice and Speech, 2012), 22–24.
- Vassilakis, Pantelis N. "Perceptual and Physical Properties of Amplitude Fluctuation and Their Musical Significance." PhD. diss., University of California, Los Angeles, 2001.
- Vennard, William. *Singing: The Mechanism and the Technic* (New York: Carl Fischer, 1967).
- Voice Production: The Vibrating Larynx.* DVD. Janwillem van den Berg and William Vennard, 1960 (University Park: Pennsylvania State University, 2013).
- "Voice Qualities." The National Center for Voice and Speech.
<https://ncvs.org/voice-qualities/>
- von Wesendonk, K. "Ueber die Synthese der Vokale aus einfachen Tönen und die Theorien von Helmholtz und Grassmann." *Physikalische Zeitschrift* (1909).
- Watson, Peter J., and Thomas J. Hixon. "Respiratory Kinematics in Classical (Opera) Singers." *Journal of Speech, Language, and Hearing Research* 28, no. 1 (March 1985)
<https://doi.org/10.1044/jshr.2801.104>
- Watson, Peter J., Thomas J. Hixon, Elaine T. Stathopoulos, and Daniel R. Sullivan. "Respiratory Kinematics in Female Classical Singers." *Journal of Voice* 4, no. 2 (January 1990).
[https://doi.org/10.1016/S0892-1997\(05\)80136-5](https://doi.org/10.1016/S0892-1997(05)80136-5)
- Weenink, David J. M. "Formant Analysis of Dutch Vowels from 10 Children." *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 9 (1985).
- Weiss, A. P. "The Vowel Character of Fork Tones." *The American Journal of Psychology* 31, no. 2 (April 1920).
- Welch, Graham F., David M. Howard, and John Nix, eds. *The Oxford Handbook of Singing* (Oxford, UK: Oxford University Press, 2019).

- Winckel, Fritz. *Music, Sound and Sensation: A Modern Exposition* (New York: Dover, 1967).
- Zhang, Zhaoyan. "Mechanics of Human Voice Production and Control" *Journal of the Acoustical Society of America* 140, no. 4 (October 2016). <https://doi.org/10.1121/1.4964509>
- Zwislocki, Jozef J. *Auditory Sound Transmission: An Autobiographical Perspective*. (New York: Psychology Press, 2013).

Index

Page references for figures are italicized.

- abduct, 190–91, 202
- absolute spectral tone color (ASTC), 129–36, 138–41, 147, 149, 152, 155–59, 161, 163, 167–69
 - definition, 131
 - labels, 132–33, 140–41
 - limited subset of timbre, 138
- acoustics, XIII, 3, 5, 9, 11, 16–21, 23, 25, 28, 30–31, 38, 55, 61–62, 67, 70–71, 78, 87, 100, 104, 106, 119–21, 125, 129, 151, 160, 163, 168, 177, 181, 183–84, 190, 194, 196–97, 207–8, 215;
 - acoustic registers, XIII, 73, 205
 - acoustic theory of voice production, XIII, 120
 - voice acoustics, XII, 15, 22–23, 33, 87–88, 211
- adduct, 104–5, 107, 161, 186–87, 190–93, 196, 198, 204, 210
- adjacent functions. *See* function
- air mass, 18–20, 69–70, 76–78, 87, 103, 106, 111, 150, 190, 193–98, 216
- airflow, XIV, 18–19, 21, 41, 49, 68, 83, 103–04, 107, 161, 190–94, 196, 198, 201–4, 210–11, 215–16;
 - transglottal. *See* flow
- amplify, 21, 25–26, 30, 54–55, 58, 60, 62, 70, 76, 78, 81, 111–12, 135, 202, 207, 209
- amplitude, 12, 15–16, 20, 26, 29, 34, 70, 73, 76–82, 87, 101, 112–13, 116, 135–36, 138, 147, 150–52, 155, 160, 165, 168, 170–72, 197–98, 202, 208–09, 216, 220–21
- Angus, Jamie, 65n14, 93–93, 98
- anthropomorphize, 39, 134, 140
- the anxiety of influence, 8
- ASTC. *See* absolute spectral tonal color
- asymmetrical, 21, 83, 162, 191, 194, 197–98

attack, 4, 67, 70
attention, 13, 30, 45, 111, 118, 130–31, 152, 193
attenuate, 20, 41, 72, 111–112
auditory:
 attribute, 45
 canal. *See* ear
 cortex, 47–48, 79, 124
 perception, 60
 roughness, 3, 29–30, 42, 62–63, 91–93, 95–101, 104–8, 121, 125–26, 140, 143–44, 146–49, 153, 156–60, 163, 169, 171–72, 181–82, 192–93, 205–6, 208, 215, 223–24
 scene, 30
 sensation, 36
 transduction, 31–32, 50, 88, 153, 234

balloon, 19–20, 69–71
bandwidth, 12, 99;
 equivalent rectangular, 97–98, 108, 173, 215
 formant, of, 159
 telephone, analog, 11, 12, 57, 68, 223
beneficial narrowing, 100, 116, 195–96, 199, 202–4, 210–11
Bareilles, Sara, 166–68, 170, 195
beats, 93–96, 223;
 binaural, 95
belt, 5, 78, 105, 164, 166, 168, 205–6, 211
Bloom, Harold, 8–9
boundary, 27, 70, 132, 183
Borch, Daniel Zanger, 86
Bozeman, Kenneth, XII, 47, 64n5, 73, 76–77, 89n22, 108n7, 117, 151, 206
Bregman, Albert S., 37
bright(-ness) 55, 157, 161, 168, 186, 193–96, 202–9, 215–16;
 auditory roughness, and, 41–42, 48, 100, 106, 121, 163, 170, 192, 205
 chiaroscuro, 114
 dynamics, and, 105, 210
 formants, and, 118, 208
 frequency, and, 113, 117, 126, 129–30
 genre, and, 207
 hearing sensitivity, and, 171
 history of, 118–19
 mix, and, 104
 narrowing in vocal tract, and, 195, 202, 204
 pitch, and, 49, 55, 58, 106, 114, 121

ringing, 54
singer's formant cluster, and, 54
stochastic turbulence, of, 103
tone color, 4, 10, 33, 47, 108, 111, 113–115, 117, 125, 131, 134, 143, 156
vowel identification, and, 60, 100, 123, 126, 133, 140
broadband, 19–21, 34, 71, 73, 194
Brown, James, 60–61, 86
buzz(-y, -ing), 3, 10, 22–23, 29, 33, 41–42, 45, 48, 49, 52, 54–56, 61, 91–92, 95–97, 99–100, 102–7, 121, 124, 138, 143–45, 148–50, 154, 158, 163, 170–71, 173, 192, 195–96, 198, 202–5, 207–10, 215, 223

Carpenter, Alan, 118
CCM. *See* contemporary singing
charlatan(s), 7–8, 10
chest voice. *See* register
clap, 19–20
clarify, 41, 121, 178–80, 211, 217
clear, 118, 135, 151–52, 155, 172, 180, 195–96, 199, 203–4, 206–7, 211
closed quotient. *See* contact quotient
Coffin, Berton, XII, 76, 165
Cogan, Robert, IX, 37, 119–21, 131, 135
timbre quote, 37
cognition, 125, 129, 184, 233;
language cognition, 10, 58, 129, 134, 140–1
commercial recordings, issues with, 56, 165–66
complex, XI, XIII, 2, 5, 8–9, 14, 16, 18, 23, 25, 30–31, 42, 46, 49, 62, 66, 69–71, 73, 78–80, 83, 99–100, 102, 111, 113–16, 123, 125, 131, 133, 140–41, 147, 152–53, 155–56, 158–59, 164, 168, 172–73, 180, 186, 192, 205, 207–8, 220, 223;
timbre, 47, 208
tone, 32, 39
waveform/pressure pattern/sound, 14, 20–22, 28–31, 46–47, 49–53, 68, 75, 78, 81, 87, 95, 100, 111–12, 114, 117–18, 124, 130, 143, 147, 221
computer, XII, XIII, XIV, 2, 18, 33, 41, 67, 112, 120, 137, 144, 164, 184, 216
constructive interference, 20, 80, 197
contact quotient, 121
contemporary singing, 54, 56, 159–60, 164, 172–73, 206–7
continue, 179–80
Corelli, Franco, 166–67, 183
critical band of hearing, 30, 92, 96–98, 100, 147, 215
Cross, Melissa, 86

dampen, 49–50, 69–71, 76, 77–78, 81, 83–84, 86, 100, 112–113, 116, 136, 148, 150, 159–60, 163, 164, 168, 184, 195–96
dark, 4, 10, 33, 47, 114, 117–18, 121, 134–35, 143, 152, 157, 159, 170, 173, 181, 194–95, 205, 207
decay(-s), 20–21, 70–71, 73–74, 79, 170
declarative. *See* language
decomposition. *See* frequency decomposition
Denes, Peter, 116, 126n10, 128n39
density, 19, 194
destructive:
 information-destructive, 32
 interference, 20, 76, 197
diameter, 70, 198
difference tone, 94, 108
discrete, 173;
 discrete Fourier transform, *See* Fourier transform
 formant, 158, 171
 frequency components, 67
 harmonics, 19, 23, 76–77, 87
 levers, 181
 measurements, 217
 parts of the voice, 22
 percept, 52
 processes, 32, 151
 resonance, 76
 steps, 178
 subsystems, 18
displacement:
 abdomen, 188, 197
 air, 178, 194
 basilar membrane, 96
 cochlear fluid, 27
 ribcage, 188, 198
 vocal folds, 191
distortion, 60, 61, 85, 86, 191
duration, 21, 42, 51, 53, 64, 67, 69, 72, 73, 75, 83–85, 88, 155, 195, 207, 216, 221–22;
 pitch of speech, and, 114
timbre has, 73, 87, 162, 206

ear, 9, 16–18, 22–24, 31–34, 38, 42–43, 63, 72, 92, 95–96, 112, 137, 144, 171;
 apex, 26–28, 30

ear (auditory) canal, 21, 24–25, 30
base, 26–30
bony labyrinth, 25
basilar membrane, 27–30, 34, 47–48, 50, 92, 96, 98, 100, 124, 215
cochlea, 25–33, 47, 50, 79, 80, 96
cochlear duct, 27–28
cochlear fluid, 25–27, 79, 80
cochlear nerve, 28, 30,
eardrum (tympanic membrane), 21, 24–26, 30, 34, 79
endolymph, 28
footplate, 26, 28
incus, 24
inner, 21, 24–26, 30–31, 50, 53, 65n14, 79, 172, 215
malleus, 24
mediation, 23
middle, 21, 24–25, 34
organ of Corti, 28
ossicles, 24–26, 30, 79
outer, 24, 25
perilymph, 28
pinna, 24–25
Reisner's (vestibular) membrane, 29
round window, 27
scala tympani, 27–28
scala vestibuli, 27–28
stapes, 24, 26–28, 79–80
stereocilia, 28–29
ease(-fulness), 22, 41, 157
Eilish, Billie, 59, 60–62
elastic recoil, 188, 191;
 abdomen, 188–89
 air, of, 194
 ribcage, of, 187, 197
 vocal folds, of, 191–92
engage, 178–79, 188
Engel, Gustav, 118
Engelhardt, V., 118
equivalent rectangular bandwidth (ERB). *See* bandwidth
evidence-based voice pedagogy, 7

Fales, Cornelia, 131
falsetto. *See* register

filter, 20–21, 25, 28, 34, 38, 95, 157, 159, 172, 190;
pass-. See frequency

Fiuza, Mauro Barro, 86

Fletcher, Harvey:
transient versus steady-state models, comments on, 68

flow, XIV, 17–18, 31, 32, 67, 70, 100, 103, 191–95, 197–98, 199;
nasal, 196

phonation, 100, 103, 106, 192

resistance, 103, 194, 197, 210–11

transglottal, XIV, 18, 41, 49, 68, 83, 103–4, 107, 191, 194–96, 198, 215–16, 217

formants, XIII, 23, 34, 76, 84, 112, 113–18, 120–21, 123–26, 132–35, 138–40, 143–45, 147–53, 158–64, 166, 168–70, 172, 197, 205–7, 215;
history, 112

in common between two vowels, 123

plot graph, 9–10

scaling, 115

singer's formant (-cluster, SFC), 11, 54–56, 206

tone color of, 120

tracking, 153

tuning, 63, 76–77, 88, 105, 121, 124, 159

Fourier transform, XIII, 15–16, 18, 22–23, 28, 31–33, 47, 51, 67, 69, 112, 151, 161, 216, 221;
discrete (DFT), 18, 221
fast (FFT), 18, 221

frequency:
dependent tone color, 32, 39–40, 42, 52, 130–32, 135, 139–40, 144, 146–47, 156, 160, 162

decomposition, 18, 31, 50

domain, 46–47, 51, 130, 148

filter, 20, 25, 28, 34, 190

high-, XIV, 28–30, 32, 48, 51, 52, 58–60, 96, 111, 113, 117, 125, 134, 147, 170, 207

low-, 28, 30, 48, 51, 111, 113, 147

pass-filter, 38, 95, 157, 159, 172

resolution, 29–30, 92, 96, 215

sampling frequency, 217

function, 21, 162;
adjacent or intermediate, 140, 177, 201, 203–4, 209, 211

the ear, of, 23

language, in, 115, 133

listening, functional, XII, 3, 41

Praat, in, 152

time, of, 49, 53
voice (vocal), 3, 15, 26, 41, 91, 100, 106, 140, 177, 183, 196, 198, 201–2, 204, 207–8, 211
vocal tract transfer, 87
fundamental, 31, 38–39, 40–41, 43, 46–53, 55, 58, 60, 63–64, 68, 72–73, 77, 80–82, 87–88, 94, 99, 106, 111, 116, 118, 121, 122, 123, 124, 136, 143–45, 147–51, 153, 155–57, 161, 170, 172, 181, 197, 203, 206–7, 209, 220; missing, 47, 65n6, 88, 156, 216, 221, 223; obvious true, 155–56
tone color of. *See* tone color

Gautereaux, Kayla, 181
Gehrcke, E., 118
glottal, 171;
 adduction. *See* adduct
 closing (closure), 10
cycle, XIV, 21, 23, 51, 68, 72, 75, 83, 87, 100, 103–4, 112, 151, 162, 190, 192–93, 196, 202
contacting, 70, 116, 164–65
impulse, 21, 23, 31–32, 68, 70–73, 76–77, 81, 83–86, 100, 102–3, 107, 150–51, 160, 162, 165, 168, 170, 194, 197–98, 220
oscillation, 168
resistance, 103, 190
sub-. *See* pressure
supra-, 18, 106, 190, 192
Grassman, Hermann, 118

handclap. *See* clap
Hampson, Thomas, 167–68, 170, 195
harmonic(s), 5, 22–23, 31–32, 46, 48–50, 52, 53–54, 57, 57, 60, 64, 68, 71–73, 76–77, 80–81, 84, 86–88, 91, 93–97, 99–103, 106–107, 111–113, 116, 119, 122, 131–32, 135, 144–45, 147–49, 151–56, 158–60, 162–63, 166, 168–70, 183, 197, 205–7, 216, 221–23;
generated by skewed airflow, 6
generated by vocal folds, 5, 15, 19, 23, 34, 76, 112
model, 22, 96
resolved, 48, 54, 55–56, 58, 121, 223
series, 55, 81, 91, 93, 97, 101, 148, 153, 159–60
spectrum of, XIV, 15–16, 21–22–23, 47–48, 51, 53, 55, 67, 72, 87, 147
sub-, 77, 82–85, 88, 197
theory of voice production, 67
unresolved, 48–49, 54, 56–58, 153, 223

head voice. *See register*

hearing, 16, 23, 24, 29, 30, 33, 42, 60, 91, 85, 102, 124–25, 169, 176, 223; acuity, 62
aids, 60
loss, 60
other animals, 134
range of, 24
sensitivity, 25, 28–28, 30, 34, 162, 171, 208

Heller, Eric J., 21, 92, 94

Hermann, Ludimar, 112

Herbst, Christian, 102

hertz (Hz), 49, 113

heuristic(s), 4, 13, 16, 22, 198, 213

Howard, David, 92–93, 98

Howell, Ian, 32, 130–32; how do I sing high notes loud quote, 183
vowels are tone color-like quote, 119

impedance, 26

impulse, 19–20, 172; electrical, 25, 30
glottal. *See glottal*
source. *See source*
nerve, 31, 124

inertia, 5, 18

inharmonic theory of voice production. *See transient*

intensity, 4, 10–11, 20, 23, 26, 30, 33–34, 38, 45, 58, 73, 87, 95, 97, 100, 105–6, 130–32, 135, 140–41, 144, 149, 152–53, 156, 161, 163–64, 166, 170, 180–81, 197, 202, 221–23, 225

interdisciplinary work, XI, 17

intermediate sounds. *See function*

international phonetics alphabet (IPA), 9–10, 11, 133, 134, 141

Köhler, Wolfgang, 118

Konno, Kimio, 189

label, 5, 10, 33, 37, 60, 76, 91, 99, 111, 131–33, 135, 141, 159

Lamperti, Giovanni Battista, 9

language, 10, 15, 57, 114–16, 129, 133, 210; declarative, 184, 215
procedural, 176–77, 179, 184, 216
teaching, of, 176
visual, 16, 67

language cognition. *See* cognition.

larynx, 54, 116, 186–87, 193–94, 203, 209;
arytenoid cartilages, 102, 106, 181, 193, 211, 215
cadaver, 23
creating harmonics. *See* harmonics
cricoid cartilage, 102, 187, 203, 215
cricothyroid (CT), 102, 187, 192, 215
epiglottis, 196
false (vestibular) folds, 191
interarytenoids (IA), 102
lateral cricoarytenoids (LCA), 102
posterior cricoarytenoids (PCA), 102
thyroarytenoid (TA), 102, 103, 105–6, 180–81, 192–93, 204, 216
thyroid cartilage, 102, 186–87, 203, 215
vocal folds, 2, 5, 9, 17, 18–23, 32, 34, 51, 60, 63, 68–69, 72, 76, 77, 82–83, 86–87, 100, 102–3, 105–6, 111, 116, 117, 151, 162, 164, 180–81, 185–88, 190–98, 204, 208–10, 215–16
vocal folds body, 102
vocal folds cover, 162, 181, 192, 202
vocal folds depth of contact, 100, 181, 191–93, 211
vocal ligament, 181, 192
linear, 175, 185, 192, 221
listening functionally. *See* function
loudness, 4, 9, 37–38, 41–42, 123, 130, 164, 166, 181–83, 191, 202, 208, 210–11, 216, 223;
equal-loudness contours, 25–26, 152

Mach, Ernst, 117–18
masking, 29, 54
McAdams, Stephen, 62
 there does not exist quote, 38
McCoy, Scott, 45, 72;
 a subjective perception of pitch McCoy quote, 45
Mead, Jere, 189
mechanical:
 action, 55
 motion, 24
 sound, 141
 vibration, 25–26
Merman, Ethel, 168–70
Merriman, Dan, 177
microphone, 20, 62, 138, 166, 220

migrate/migration. *See vowel*

Miller, Donald G., IX, XII, 76–77, 84, 102, 112, 120–21, 124, 168

Miller, Richard, XII, 156–57

mix. *See register*

mix-belt. *See register*

models, XIII–XIV, 1, 3–4, 6–10, 11, 13, 15–18, 22–25, 28, 30, 32–33, 42–43, 49, 55, 59–60, 62, 67–69, 76–77, 86–88, 100, 112, 115, 117, 129, 135, 158–59, 184–85, 193, 201, 233;

- harmonic model. *See harmonic*
- nuance surrounding, 5, 118
- problems with, XIII–XIV, 21
- singer’s formant (-cluster, SFC) model, 58
- source-filter model. *See source-filter theory*
- steady-state model. *See steady-state*
- time-and-pressure-domain model, 31, 69, 79, 86–88, 111–12, 116, 206–7
- transient model. *See transient*
- two-timbre model, 102, 121, 125, 171, 207

modification. *See vowel*

Morton, John, 118

musical theater, 164–65, 196

narrowing, 100, 182, 195–96, 199, 202, 204, 207;

- anterior to posterior, 196
- beneficial or useful, 12, 195, 203, 210
- lateral to medial, 196

nasal. *See flow*

noise, 3, 10, 33, 46, 48–49, 52, 54–55, 58, 61–63, 86, 91, 99, 106–7, 131, 148, 150, 154, 156, 170–71, 172, 181, 192–93, 195–96, 208–9

nonlinear, 69, 77, 93, 98, 160, 164, 185, 198, 204, 209

nose, 19, 196

Obert, Kerrie, 196

onset, 41, 67, 193, 223

oscillate, 2, 5, 18, 20–21, 23, 29, 50–51, 60, 68–72, 76–77, 79–83, 85–87, 91, 93, 99–100, 111–12, 116–17, 121, 136, 143, 147–48, 150–153, 155–56, 159–60, 162–65, 167–71, 181, 185, 190–92, 195, 197–98, 202, 208, 216

overtone, 80–81, 114, 123, 160;

- singing, 63, 79, 81–83, 87–88, 97, 155, 159–60, 195, 207, 223

Pasquale, Steven, 168, 169–170, 195

Patterson, Roy D.:

- wastebasket quote, 37

percept, 51, 87, 100, 106, 124, 151, 153, 156, 158, 160, 163, 206–7, 216, 223–24;

auditory roughness, of. *See* auditory brightness. *See* bright
buzzy. *See* buzz(-y, -ing)
fundamental, of, 40, 52
fused, 92
listener's, 18, 38, 114, 125
noisy, 86
pitch. *See* pitch
pure, 53, 161, 163
slippery, timbre is a, 130
sound, of, 32
spectral centroid, of, 153
timbral, 40
tone-color. *See* tone color
vowel, 133, 140, 156–57, 160–61
perception, XII–XV, 1–3, 5–6, 9, 13, 16–18, 31, 33, 37, 42, 45, 47–49, 56, 60, 62, 80–81, 83, 88, 93–94, 97, 99, 119, 124–25, 130–31, 140, 170, 17–77, 184, 198, 216
period, 21, 29, 47–48, 49–50, 64, 68, 71–73, 75, 79–83, 87–88, 100, 116, 118, 150–51, 155–56, 160, 162–63, 165, 167–69, 197, 207, 216, 220;
pitch period, 51, 69, 72–73, 75, 77, 79, 81, 83, 87, 91, 96, 116, 136, 148, 156, 159, 161, 163–64, 165, 167–70, 197, 220
sampling period, 219
periodic, 15, 16, 22, 23, 29, 31, 47–49, 51–52, 60, 67–73, 79, 82, 85–87, 90n26, 93–94, 100, 106, 112, 131, 144–45, 148, 151, 208, 216, 220, 223;
quasi-, 46
Perna, Nicholas, 169
pharynx, 19, 181, 186, 195–96, 198–99, 202
Pinson, Elliot N., 116
piston, 26, 79
pitch, 3–4, 5, 9, 10, 11, 15, 21, 29, 32–33, 37–39, 42, 45–64, 67, 69–70, 72–73, 75, 77, 79, 82–86, 88, 91, 94–95, 97, 100–2, 104–8, 113–14, 116–20, 123–24, 129–30, 135–37, 139–41, 143–44, 146–51, 153–57, 159–64, 170–172, 180–83, 186–87, 191–93, 197, 202–11, 216, 223–24;
noise, of, 60–61, 86
percept, 46–52, 53, 58, 60, 61, 63–64, 72, 80, 82–83, 86, 94–95, 101, 130–31, 146, 148, 156, 160, 172
period. *See* period
place theory, 47–48, 50, 64
resolution/resolved, 55, 91, 100–1, 106–7, 143–44, 146–49, 153, 157–58, 171–72, 226
speech-pitch range, 58, 79, 104, 135, 140, 164, 168, 183, 196

subjective perception McCoy quote, a. *See* McCoy, Scott
threshold, 49, 50, 70, 93, 95, 99, 144, 148
timing theory, 47–48, 50, 64, 80, 83, 88
unresolved, 50, 55, 58, 63, 101, 106, 121, 138, 144, 154
Plomp, Reinier, 39, 46, 49, 118–19, 130–31;
 simple tones have timbre quote, 39, 118, 130
pop, 19–20, 69–71, 84, 144, 196, 223
Praat, 38–39, 95, 131, 144, 152, 220
practical, XIV, 7, 12, 16, 38, 40, 55, 72, 92, 97, 100, 102, 112, 182, 195;
 application, 8, 16, 33, 130, 176, 182
dismissal unless immediately, 176
pedagogical framework, 4, 175
teaching models, 16
the voice studio, in, 5
pressure, XIV, 17, 20, 28, 31, 37, 50, 73, 107, 150, 168, 191, 194, 197, 216–19;
 atmospheric, 19, 69, 75, 220
 between the folds (intraglottal), 18, 190–91, 193–94
 change, 20, 24–25, 28, 29, 31, 34, 46, 51–52, 67, 69, 84, 100, 106–8, 111, 141,
 143, 145, 190, 194, 194, 220
 cycle, 150–51, 156, 161
 complex pattern, 15, 21–22, 30–31, 52, 67–68, 70
 curve, 19
 drop, 19–20, 29, 70, 77, 79, 83, 100, 103, 106, 151, 162, 164, 191, 197
 dynamics, 18
 high, 18–19, 75, 164, 194, 196, 198
 impulse. *See* glottal
 low, 75, 150, 197, 198
 maximum, 29
 oscillation of, 5, 20, 51, 77, 79, 100, 111–12, 171
 pattern, 29, 31, 47, 50–51, 70, 72–73, 75, 88, 95, 121, 136, 145, 147, 153, 163,
 198–99, 221–22
 phonation threshold, 181, 190–91
 processing by the cochlea, 33
 rise, 29
 skewed pattern, 6
 supraglottal, 18–19, 68, 70, 76, 164, 190, 192, 196, 198, 210
 time-and-pressure-domain model. *See* models
 tracheal, 18, 103, 105, 171, 175, 181, 187–92, 198, 203–4, 210
 waves, 18–20, 22–25, 27–28, 30–34, 46–47, 50, 69–70, 73, 75, 79–82, 123, 138,
 168, 181, 206
procedural. *See* language
proprioception, 177

psychoacoustics, XIV, 4, 16, 30, 34, 46, 92, 131, 155, 157, 180, 201, 213, 216
pure, 3, 10, 33, 51, 56, 79, 81, 91–92, 96, 101–6, 117, 119, 124, 138, 155, 159–61,
163, 171, 199, 202–3, 205, 207–8;
tone, 5–6, 22, 28–29, 87–88, 92–93, 94–96, 99–100, 107–8, 111, 113–14, 117,
119, 123, 125, 129–33, 135, 138–41, 143–44, 145, 147, 149–50, 152, 156,
159, 164, 196, 207, 216, 223–24
PVC pipe, 20

quasi-periodic. *See* periodic

reflection(s), 20, 69–70
refraction, 69
register, 108, 120, 183, 202, 205, 210, 216;
 acoustic, XIII, 2, 73
 broken, 182, 205
 challenges to discussing, 102
 chest, 102–5, 181, 205, 208, 211
 exist when loud enough, 209
 falsetto, 102
 head, 103, 180, 205, 211
 laryngeal, XIII, 2, 102
 mix, 104, 182, 196, 205, 208, 211
 mix-belt, 105, 164, 166, 205
 seamless transitions, 104
 there are no registers, 201, 204
 two defined registers, 211
registration, 2, 10, 33, 41, 64, 102, 107, 116, 123, 129, 183, 191, 216;
 acoustic registration transitions, 156–7, 205
 challenges to discussing registration, 102
 heavier registration and auditory roughness, 100, 192
 mix. *See* register
 mix-belt. *See* register
 models tied to physical structures, 193
 neutral registration, 195
 on and above treble staff, 155, 206, 207
 perceptual qualities of, 205
 rebalancing, 203, 205
 triangle, 180–82, 208–9
Reid, Cornelius, XII, 177, 180–81
release, 178–80, 185;
 elastic recoil, of, 191, 204
cracks when muscles suddenly, 193

resistance:

- abdominal, 186
- airflow. *See* flow
- airmass, of, 194
- cochlear fluid, 26
- futile, is. *See* TNG
- tone color model, to, 114
- vocal folds, of, 190
- vocal tract, of, 196–97

resolved. *See* harmonics

- resonance, 88, 135, 144, 150, 153, 160, 163, 184, 216, 226;
 - amplitude of, 70, 116, 155, 165, 171, 198, 208
- auditory canal, of, 25, 30
- averaging in spectrum, 112
- basilar membrane, of, 29–30, 96
- brightness, and, 113
- damping, 69, 81–82, 87, 116, 150, 159, 163, 165, 168, 170, 184
- excitation by flicking throat or whisper, 172
- faster than vowel, 170
- first resonance, 38, 48, 112, 121, 157, 161, 165, 206, 208
- formant, relationship with, 76, 112, 152, 215
- function of time within a glottal cycle, as a, 21, 49, 53, 70, 72–73, 75, 75, 77, 79–80, 87, 111, 148, 150, 151, 155, 164
- hidden, 152
- impulse and decay, pattern of, 21, 49, 73, 77, 87, 165, 197, 208
- independent, 21–22
- interaction with source harmonics, 21, 77, 112, 122, 151
- interaction with subsequent glottal contacting, 77, 106, 151, 164, 170, 197–98
- noise, excited by, 60–61, 86
- ossicles, of, 25, 30
- oscillating pressure patterns, 19, 20, 73, 81, 111, 148, 164–65, 169, 197
- overlapping resonant responses in voice production, 69, 71–72
- overtone. *See* overtone
- pitch percept of, 63, 148, 172
- precede harmonics temporally, 72, 86
- second harmonic, interactions with, 121
- second resonance, 168, 183–84
- sensation, 9
- source-filter model technically wrong, Heller, 21
- source-filter model, within, 16, 76, 112–13, 113, 116, 121
- speed of, 72–73, 75
- strategies, 159–60, 164, 166, 183, 195

timbral information, 51, 71, 75, 150, 151
tongue, 195–96
transient theory model, within, 76, 84, 113
tube, 164
tuning, XIII, 76, 156, 161, 162, 183
vocal fry, 84, 85
vocal tract, no one place in, 20, 34
the vocal tract, of, 5, 18, 20, 21, 25, 116, 145, 149, 154, 164
vowel, 76, 171
resonator, 19, 22, 34, 54, 196–97
response:
 air, of, 194
 basilar membrane, of, 28–29, 92, 96, 98, 100
 critical bands, of, 30
 larynx, of, 186–87
 microphone, frequency response of, 166
 perceptual, 119
 small speakers, of, 48
 sound, to, 18
 speed of pressure changes, to, 111, 121
 stimulus and, 177–83, 185, 21
 tone color, 123–24, 138
 tympanic membrane, of, 25
 unified expressive response to a thought, 175
 vocal tract, of, 19, 31–32, 69–73, 76–77, 84, 87, 111, 116, 163, 172, 190, 198
reverberant, 20, 69–70
roughness. *See* auditory

sawtooth, 22–23, 95, 98
science, 8;
 acoustics, of, 9
 applied to art Vennard quote, 7
 -based pedagogic models, 17
 -curious, XI
 -derived, 8, 13
 -informed, 22
 practice, and, XII
 speech, 119
 voice, XIV, 67, 134, 215, 219–20, 183, 223
Seidenburg, Kai, 62
 there does not exist quote, 38
sensation, 9;

auditory, 37
register break, of, 205
sensitive, 25, 28–28, 30, 34, 139, 162, 171, 208
signal chain, 21, 32, 60–62, 207
sine tone, 5–6, 31, 120, 131, 141, 144, 155, 216
singer's formant cluster. *See* formants
sinusoidal, 6, 46, 80, 119, 193, 216
skew, 6, 73, 193, 198, 208–9
Slawson, Wayne, 39, 119
source:
 harmonics. *See* harmonics
 impulse, broadband, 20–21, 32, 77, 194, 197–98
 impulse as filter, 21
 limited range and capability in speech Titze quote, 116
 sound source, 21–22, 32, 87–88
 supraglottal pressure change per period, is, 190
 tone color regardless of, 39, 113, 121
source-filter theory, XIII, 5, 15, 19–20, 22, 76, 111–12, 116, 135, 168
 independence of (Titze), 116
spectral:
 analysis, value of, 31, 120
 aspects of changes in loudness, 41
 average, 113
 auditory roughness, signs of. *See* auditory
 centroid, 152–53
 components, 22
 details of radiated sound, 111, 144, 170
 domain, 51
 features do not automatically indicate percept, 107, 143
 features, opposable, 143, 159
 features point to oscillations, 112, 136, 143
 fluctuations in speech, temporal, 140–41, 160
 formants, historical role in, 112, 215
 fundamental, 40, 64, 172
 images, 18
 peaks. *See* formants
 ranges related to pitch percept. *See* pitch
 slope, 102, 107
 structures, 106
 terms, voice production framed in, 21
 tone color of peaks, 132, 135–36, 141, 151
 vowels share spectral peaks, 114, 121–22

spectrum, 225;
asymmetrical change with dynamics, 105, 125
audibility of details, 17, 147
auditory roughness, locating, *See* auditory
basilar membrane filtering analogy, 25, 28
capping frequency range, issues with, 58, 60
distinct qualities across, 41, 43, 46–47, 91, 106, 117, 122, 124, 151, 156, 160,
171
explanation, XV, 219, 221–22
Fourier transform output, 16, 18, 22–23, 31, 68, 87, 116
fundamental. *See* spectral
issues connecting output to voice production, 15, 18, 32–33, 67, 69, 99, 153
long term average spectrum (LTAS), 54, 57, 122
pass-filtering. *See* frequency
pitch, locating. *See* pitch
proximity effect of microphone. *See* microphone
tone color/are vowels encoded across entire, 5, 40, 132, 133–34, 140, 148, 155
windowing, 31–32, 69, 87, 105, 221–22
spectrogram, XIV–XV, 5, 15, 17, 18, 28, 31–32, 36, 40, 41, 46, 48, 49, 50, 51, 52,
63, 64, 67–69, 71, 83, 87–88, 96–97, 99, 102, 107, 116–17, 124, 126, 143, 147–
48, 152–53, 159, 162, 164, 219;
explanation, 222–23
spectrograph, 18, 38, 57
speech:
coarticulation, 134
formants of. *See* formants
historical frequency range, 56–57
intelligibility, 12, 123, 129, 225
measurements of, 33
phonemes in, 11
pitch and intensity, 4, 58, 79, 104, 114, 116, 135, 164, 168
production, mechanisms of, 68
research, 9, 112, 115, 119
semantics, 138
telephone, over. *See* telephone
tone color, resemblance of, 39, 118, 140
-wave has few segments Denes and Pinson quote, 116
steady-state, 67–68, 77, 88
stiff(-ness):
air, 194, 196, 198
basilar membrane, 28
fibers of thyroarytenoids, 181, 192

ribcage, 178

stimulus:

basilar membrane, 29, 47–48, 96, 100

breathy sound, 193

brightness, 113

complex pattern, 112

critical bands, within, 30, 92, 100, 215

cochlea, 25, 30, 32

distance from source, and, 171

the ear, 17, 23, 30, 72, 143

ear, different sources stimulate similarly, 33

functions indirectly, 198

onset, glottal, 193

pitch, high, 153, 183, 192, 206

pitch, low, 191–92

pitch processing after inner ear, 50

response of the body, organized, 175, 177–82

simple idea, 184

sound you want to make, clarify, 179, 205, 209, 211

stereocilia, 29

timbre and pitch, 69, 130

tone color, 119

Story, Brad, 21

Stumpf, Carl, 118

subsystems of voice:

articulation, 16, 184, 196

perception, XIII, 16, 62, 184

phonation, 16–19, 70, 82–85, 100, 103, 105–7, 160, 164, 171, 172, 184–85, 188, 190, 192, 194, 197, 208, 210

radiation, 16, 62, 184

resonation, 16, 177, 185, 190, 216

respiration, 16, 178, 184–86, 189–90, 193

Sundberg, Johan, 104

roughness of the timbre, 93, 101

singer's formant (-cluster, SFC), and, 54

voice source qualities, 102

superimpose, 73

Švec, Jan, 83, 102

swing:

displaced air molecules, 194

pushing analogy from Titze. See Titze, Ingo R.

synthesize, 31, 146, 176, 225

Tallon, Tina, 12

Telephone:

analog, 9, 47, 57

bandwidth. *See bandwidth*

Tempesta, Mark, 169

timbre, 11, 67, 225

absolute, 130

auditory roughness of, 93

brightness, and. *See bright*

comparing voice and instrument, 40

complexity of, 42, 147

compositional parameter of music, 120

control factor for registration, 182–83, 187, 191, 202–3, 208–11, 216

close timbre, 206

definitions, 9, 37–38

duration of. *See time*

dynamics, relationship with, 105, 210

grave, neutral, and acute, 120, 131, 135

high-frequency components of periodic sound, 60

impulse, of, 20, 72

inherent, 123

intelligibility without second formant, 134

-invariant aspects of timbre, 41–42, 130, 133

musical character, 11

pitch, loudness, and, 4, 33

pitch, relationship with, 46–47, 58, 62–63, 68, 144, 155, 157

parsing qualitative aspects of, 1, 17, 25, 32, 37, 42

pure tones, of, 39, 119

similar qualities between different sounds, 38

singing voice, 11, 21, 23, 32, 37, 114

slippery percept, is a Fales quote. *See Fales, Cornelia*

spectrogram, in, 31, 48–49

timing (duration) of. *See time*

tone color, and, 111, 113, 116, 130, 156

two-timbre model. *See models*

-variant aspects of timbre, 41–42, 133, 147

vocal fry, and, 83–85

vowel, and, 115, 181

wastebasket Patterson quote, 37

waveform, understood through, 69

time-invariant. *See timbre*

time-variant. *See timbre*

time:

aperiodicity of vocal fry, 83, 85
beats, rate of, *See* beats
damping of vocal tract over. *See* resonance
domain, 72
flow glottogram, 18
fry scream, timescale of, 86
harmonics exist at all time scales, idea that, 21, 72, 87, 116
hertz (repetitions per second), 51, 113
interaction of resonance with subsequent glottal contacting. *See* resonance
-invariant aspects of timbre. *See* timbre
lag of each subsystem, time, 185
mistiming resonance and vocal fold oscillation, 150–51, 164
play and experiment, to, 164
real-time visual feedback, 102, 124
resonance oscillates as a function of. *See* resonance
series (graph or digital file), 18, 219
space over time, moves through, 22
spectrum shows snapshot of (windowing), 31–32, 69, 87, 105, 232–22
-variant aspects of timbre. *See* timbre
timbre, and duration of, 41, 72–73, 75, 87, 116, 150, 162
time-and-pressure-domain model. *See* models
think specific thoughts at specific times, 178–79
try an exercise several times, 182
voice production, XIV,
waveform charts pressure changes over, 49, 53, 68, 75, 219–20
Titze, Ingo R., 5, 76, 105, 116, 150, 185;
pushing swing analogy, 77, 79, 150–5
source-filter independence, 116
tone color, 123–25, 157–58, 216, 225–26
1fo, of, 160, 181, 203, 205–7, 209
absolute spectral. *See* absolute spectral tone color.
brightness, and, 111, 14
complementary, 123, 132–33, 141, 163, 205
frequency-related properties. *See* frequency
formants, of, 120–22, 134, 147–53, 158–61, 163–64, 169, 171–72, 197
in common between two vowels, 121
fundamental, of, 51, 155
language, and, 116, 129
mechanism, underlying, 124
obvious true fundamental, of, 156
pitch, auditory roughness, and, 3, 32–33, 42, 62–63, 143, 172

pure tones, of, 130, 139
spectrogram, on, 126
vowels are complex combinations of, 138
vowels are tone color-like quote, 119
vowel's dominant tone color. *See vowel*
vowel-like, 119, 133

tongue:

[a] at bottom of treble staff, and, 203
behavior of tongue hidden, 184
dull and clear sounds, 195
firming, 195, 199
firming the tissue of in overtone singing, 81, 195
[i] and [e] toward top of treble staff, and, 206
moving back in mouth, 181
narrowing with, 195, 202
open vowels, and, 140
pushing down, 199
relaxed, 195
removing as an agent of narrowing, 202
vallecula, closing, 196, 199
wawa pedal exercise. *See wawa pedal exercise*
weird R exercise. *See weird R exercise*

tonotopic, 28, 124

tracheal:

pressure. *See pressure*
pull, 186–87

transfer function, 87

transglottal. *See flow*

transient:

definition, 67
impulse when vocal folds contact, 83
resonance within transient theory. *See resonance*
steady-state theory, compared to, 67–68
timing theory of pitch perception, 47
theory of voice production, 67–68
voice science, in, 67
a waveform, in, 69

unamplified, 58, 62, 138, 207–8

unresolved. *See harmonics*

Vardaman, Lynne, IX, 177–78, 180

Vennard, William, 7, 10, 114, 119, 123, 225
charlatans, on, 7
vowels share spectral peaks quote, 114

vocal function. *See* function

vocal folds. *See* larynx

vocal pedagogy. *See* voice pedagogy

vocal tract, 5, 11, 15, 18–23, 31–32, 34, 47–48, 50, 68–71, 76–77, 83–87, 106–7, 111–13, 115–16, 123, 129, 136, 139, 144–45, 148, 150–51, 156, 159–61, 163–65, 170, 172, 185, 191–98, 208, 215–16, 226;
narrowing in. *See* narrowing

overtone singing, posture for, 81

resonances. *See* resonance

source of voice, is, 32, 77, 190, 194, 197–98

vocal fold oscillation without, 23

VoceVista Video Pro, 38–39, 48, 95, 113, 144, 222

vocology:

- animal perception of animal vocalization, 134
- class, XV, 176
- community, XIV, 5
- curriculum, 3, 184
- literature, XII, XIII, 2–3, 5, 8, 34

voice acoustics. *See* acoustics

voice function. *See* function

voice pedagogy, 3;

- acoustics in, 21–22, 32–33, 87, 121
- class, XV, 176
- community, XIV, 5
- curriculum, 3, 184
- dissemination of, 8
- evidence-based, 7
- formants in, 114–15, 117, 123
- frequency in, 113
- graphs in, 219–23
- helpful information, 16
- high-frequency energy, 58
- historical, 73
- how do I sing high notes loud quote, 183
- integrating scientific information into, 2, 10, 13, 17, 34
- linguistics, 10
- literature, XII–XIII, 2–3, 5, 8, 34
- perception in, 9
- pitch in, 45, 63

registers and registration in, 107
research methods, limitations in, 64
steady-state model in. *See* steady-state
texts foreground facts about the body, 9
timbre in, 38
tone color missing in, 119, 125
using this book in a class, 176
vowel, 225;
 acoustic theory of, XIII
 adjusting problematic, 129
 all of those formants at once quote, 115
 are something you do quote, 179
 ASTC labels, 132–33, 140–41
 belt and mix-belt, favorable, 105
 clarify the, 179
 clear versus muted or neutral, 151, 153, 168, 197
 coarticulation, and, 116
 control factor, 180
 damping, differences in, 136
 encoded across spectrum, 5
 formants. *See* formants
 identification of, 115, 116, 206
 inherent pitches in a, 114
 International Phonetics Alphabet, 9–10
 migration, passive, 47, 73
 middle-voice slide, adjustment in, 202
 modification, 11, 47, 139
 over-, 117
 phoneme, as, 115
 quadrilateral, 9
 simplification as pitch rises, 155–60, 161, 163–64, 182, 206, 208
 tone colors in common, 121, 136–38, 141
 tone color of, dominant, 5, 118–20, 131–34, 140–41, 154, 163, 171, 205, 207
 tongue-related clarity, 195, 99
 under-, 117
 vocal fry conveys vowel timbre, 85
 vowel-like, 3, 33, 119, 133
 vowels share spectral peaks Vennard quote, 114

warm:

1fo, and, 124, 160, 206–8
airflow, and, 198

amplified singers, and 207
breathy sounds, 209
bright and buzzy, and 41–42
effect of equal loudness contours, 171
first formant, 118
gauzy and neutral sound, 204
lower larynx, 186
part of the spectrum, 105, 202, 208
rebalancing registration, 203
singing for brightness, and, 208
SOVT, and, 195
ways of singing, 117
vowel, 126, 128

wave:
-form, XV, 4, 15, 32, 49, 50, 52, 53, 64, 67–71, 73, 75, 79–80, 82–84, 86–88,
105, 125, 135–36, 155, 161–66, 216–20, 223
-form explanation, 219–20
standing, 70, 151, 160

wawa pedal exercise, 196, 204
Weenink, David J. M., 138–39
Weiss, A.P., 118
Winckle, Fritz, 119
weird R exercise, 196, 204
Wesendonk, K. von, 118

About the Author

Dr. Ian Howell is the founder of and chief educator at the Embodied Music Lab. He has held classroom and studio teaching appointments at the New England Conservatory of Music, the Cleveland Institute of Music, Yale College, Swarthmore, and Rutgers. He has sung in most major concert halls across America, Europe, Canada, and Japan as a soloist and with numerous professional ensembles. He has presented original research on performing arts biodynamics at the National Association of Teachers of Singing (NATS), the Pan American Vocology Association (PAVA), the Voice Foundation, the Audio Engineering Society, and the Society for Music Perception and Cognition, and has peer reviewed for Oxford University Press, the International Physiology & Acoustics of Singing Conference (PAS7+), *Musicae Scientiae*, and PAVA. Ian has been an invited guest speaker and clinician for the NATS Chat series, the New York Singing Teachers' Association, Opera Programs Berlin, Peabody Lunch and Learn, Mannes, University of Colorado Boulder, New York University, Boston Conservatory, and the San Francisco Conservatory. He is published in the *Journal of Voice*, the *Journal of Singing*, *Classical Singer*, and *VOICEPrints*, and his first book, *Advice for Young Musicians*, was published in 2023.

Ian has won professional recognition ranging from a Grammy Award and a Grammy Award nomination for his recordings with

Chanticleer to a special commendation by the American Academy of Teachers of Singing for his “work with low-latency platforms and associated technology, and broad dissemination of instruction in its use” during the COVID-19 pandemic. Ian won the Van L. Lawrence Fellowship in 2022 for work investigating cisgender male bias in a common voice science model, and he was elected to the American Academy of Teachers of Singing in 2023. His research interests include the intersection of human perception and the singing voice, with a special focus on the role of auditory transduction.

He now reaches a worldwide audience of clients and students via the high-quality, low-latency online collaboration tools he helped to curate and popularize during the Covid-19 pandemic. Ian Howell lives in Ann Arbor, Michigan with his wife and their two children. There is talk of getting a cat.