Physics of Music

Vibration, Hearing, Instruments, and the Sound of a Room

Martin Veillette

[object Object]

© 2026 Martin Veillette.

This book is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC-BY-NC-SA-4.0). You are free to share and adapt it for non-commercial purposes, with attribution, under the same licence:

https://creativecommons.org/licenses/by-nc-sa/4.0/

This book adapts openly licensed content from several sources. The per-chapter attribution ledger is `SOURCES.md`:

https://github.com/QuadriviumPress/physicsOfMusic/blob/main/SOURCES.md

The interactive simulations reproduced here as screenshots are published by PhET Interactive Simulations (University of Colorado Boulder) and by the OpenLyceum project under their own licences; each caption links to the running simulation.

Contents

Preface **vii**
0.1 Audience and prerequisites . . . vii
0.2 Organization . . . vii
0.3 How the chapters are built . . . viii
0.4 The laboratory and the project . . . viii
0.5 A note on what is left out . . . ix

I Sound and Vibration **1**

1 Sound, Music, and Simple Harmonic Motion **2**
1.1 What Sound Is . . . 4
1.2 From Sound to Music: Four Attributes . . . 9
1.3 Simple Harmonic Motion . . . 10
1.4 Energy, Amplitude, and Damping . . . 16
1.5 Summary . . . 18
1.6 Check Your Understanding . . . 19
1.7 Conceptual Questions . . . 19
1.8 Problems . . . 20

2 Wave Motion and the Speed of Sound **26**
2.1 From a Vibration to a Wave . . . 27
2.2 Wavelength, Frequency, and Wave Speed . . . 28
2.3 What Sets the Speed of Sound . . . 32
2.4 Sound Spreading in Three Dimensions . . . 34
2.5 The Doppler Effect . . . 36
2.6 Summary . . . 39
2.7 Check Your Understanding . . . 39
2.8 Conceptual Questions . . . 40
2.9 Problems . . . 40

3 Superposition, Interference, and Standing Waves **47**
3.1 The Superposition Principle . . . 49
3.2 Interference in Space . . . 50
3.3 Reflection and the Standing Wave . . . 52
3.4 Standing Waves on a String . . . 55
3.5 Standing Waves in an Air Column . . . 58
3.6 Summary . . . 62

3.7 Check Your Understanding . . . 63
3.8 Conceptual Questions . . . 63
3.9 Problems . . . 64

4 Resonance and Normal Modes 70
4.1 Free Vibration and Natural Frequency . . . 72
4.2 Driven Oscillation and Resonance . . . 72
4.3 Damping, Bandwidth, and the Quality Factor . . . 77
4.4 Normal Modes . . . 80
4.5 Resonators in Musical Instruments . . . 82
4.6 Summary . . . 84
4.7 Check Your Understanding . . . 84
4.8 Conceptual Questions . . . 85
4.9 Problems . . . 85

II Spectrum and Timbre 91

5 Fourier Analysis, Harmonics, and Timbre 92
5.1 Fourier's Theorem . . . 93
5.2 The Harmonic Series and the Spectrum . . . 97
5.3 Spectrum and Timbre . . . 98
5.4 Beyond the Steady State . . . 101
5.5 When the Spectrum Is Not Harmonic . . . 104
5.6 Summary . . . 105
5.7 Check Your Understanding . . . 106
5.8 Conceptual Questions . . . 106
5.9 Problems . . . 107

III Hearing and Perception 114

6 The Ear and the Physiology of Hearing 115
6.1 The Outer Ear . . . 116
6.2 The Middle Ear . . . 119
6.3 The Cochlea . . . 120
6.4 From Motion to Nerve Impulse . . . 123
6.5 Limits and Damage . . . 124
6.6 Summary . . . 126
6.7 Check Your Understanding . . . 128
6.8 Conceptual Questions . . . 128
6.9 Problems . . . 129

7 Loudness, Decibels, and Equal-Loudness Contours 135
7.1 Pressure, Intensity, and the Range of Hearing . . . 136
7.2 The Decibel . . . 137
7.3 Combining and Comparing Levels . . . 139
7.4 Loudness Level: The Phon . . . 141
7.5 Loudness Itself: The Sone . . . 143

7.6 Summary . . . 145
7.7 Check Your Understanding . . . 146
7.8 Conceptual Questions . . . 146
7.9 Problems . . . 147

8 Pitch, Beats, Consonance, and Dissonance 153
8.1 Pitch and Frequency . . . 154
8.2 Beats . . . 155
8.3 The Pitch of a Complex Tone . . . 158
8.4 Consonance and Dissonance . . . 161
8.5 Combination Tones and the Limits of the Theory . . . 165
8.6 Summary . . . 165
8.7 Check Your Understanding . . . 166
8.8 Conceptual Questions . . . 166
8.9 Problems . . . 167

IV Scales and Tuning 173

9 Musical Scales and Tuning Systems 174
9.1 Intervals as Ratios . . . 176
9.2 Pythagorean Tuning . . . 178
9.3 Just Intonation . . . 179
9.4 Meantone and the Well Temperaments . . . 182
9.5 Equal Temperament . . . 183
9.6 Tuning in Practice . . . 186
9.7 Summary . . . 187
9.8 Check Your Understanding . . . 187
9.9 Conceptual Questions . . . 188
9.10 Problems . . . 188

V Musical Instruments 195

10 String Instruments: Guitar, Violin, and Piano 196
10.1 The Vibrating String . . . 198
10.2 Setting the String in Motion . . . 200
10.3 From String to Air . . . 204
10.4 The Guitar . . . 205
10.5 The Violin Family . . . 206
10.6 The Piano . . . 208
10.7 Summary . . . 210
10.8 Check Your Understanding . . . 210
10.9 Conceptual Questions . . . 211
10.10Problems . . . 211

11 Wind Instruments and Air-Column Resonance **217**
11.1 Resonances of an Air Column . . . 218
11.2 Bore Shape . . . 219
11.3 The Air Jet . . . 222
11.4 Reed Instruments . . . 222
11.5 Brass Instruments . . . 224
11.6 Changing the Length . . . 226
11.7 Summary . . . 228
11.8 Check Your Understanding . . . 229
11.9 Conceptual Questions . . . 229
11.10Problems . . . 229

12 Percussion: Membranes, Bars, and Plates **236**
12.1 Why Percussion Is Different . . . 237
12.2 Bars . . . 240
12.3 Membranes . . . 241
12.4 Plates, Shells, and Bells . . . 243
12.5 Tuning the Untunable . . . 247
12.6 Summary . . . 248
12.7 Check Your Understanding . . . 249
12.8 Conceptual Questions . . . 249
12.9 Problems . . . 249

13 The Singing Voice **256**
13.1 The Voice as Source and Filter . . . 257
13.2 The Source: The Vocal Folds . . . 258
13.3 The Filter: The Vocal Tract . . . 260
13.4 Vowels, Registers, and Range . . . 261
13.5 Being Heard . . . 264
13.6 Summary . . . 266
13.7 Check Your Understanding . . . 267
13.8 Conceptual Questions . . . 267
13.9 Problems . . . 268

VI Sound in Space and in Circuits **274**

14 Room Acoustics and Concert-Hall Design **275**
14.1 Sound in an Enclosure . . . 276
14.2 Absorption and Reverberation Time . . . 278
14.3 Reflection, Diffusion, and Modes . . . 280
14.4 What Makes a Hall Sound Good . . . 282
14.5 Halls in Practice . . . 283
14.6 Summary . . . 284
14.7 Check Your Understanding . . . 285
14.8 Conceptual Questions . . . 285
14.9 Problems . . . 286

15 Electronic and Recorded Sound **293**
15.1 Microphones and Loudspeakers . . . 295
15.2 Signals and Their Imperfections . . . 299
15.3 Digital Audio . . . 301
15.4 Perceptual Coding . . . 303
15.5 Synthesis . . . 306
15.6 Space in a Recording . . . 307
15.7 Summary . . . 308
15.8 Check Your Understanding . . . 309
15.9 Conceptual Questions . . . 309
15.10 Problems . . . 310

VII Appendices **316**

16 Teaching Guide **317**
16.1 Suggested Fifteen-Week Pacing . . . 319
16.2 Dependencies . . . 320
16.3 A Ten-Week Version . . . 320
16.4 What To Trim First . . . 320
16.5 The Laboratory and the Project . . . 321
16.6 On Assessment . . . 321

17 Laboratory Exercises **322**
17.1 Equipment . . . 322
17.2 How to Write These Up . . . 326
17.3 B.1 The Oscilloscope and the Function Generator . . . 327
17.4 B.2 The Speed of Sound . . . 327
17.5 B.3 Standing Waves on a String . . . 328
17.6 B.4 Resonance and the Quality Factor . . . 329
17.7 B.5 The Spectra of Musical Instruments . . . 329
17.8 B.6 Sound Level, Distance, and Addition . . . 330
17.9 B.7 Beats and the Just-Noticeable Difference . . . 330
17.10 B.8 Measuring a Tuning . . . 331
17.11 B.9 Air Columns and the End Correction . . . 332
17.12 B.10 Reverberation Time . . . 332

18 The Final Project **334**
18.1 What the Project Is . . . 334
18.2 Choosing a Question . . . 334
18.3 The Schedule . . . 336
18.4 Giving the Talk . . . 336
18.5 What a Good Talk Sounds Like . . . 336
18.6 Rubric . . . 337
18.7 A Note on Honesty . . . 337

19 Glossary **338**

Preface

Physics of Music is a one-semester course in how musical sound is produced, transmitted, and perceived. It takes a single thread, something vibrates, the vibration travels as a wave, the wave is heard, and follows it from a mass on a spring through to a digital recording, stopping wherever music has something to ask of physics.

The subject has an unusual advantage as a physics course: the reader already has decades of data. Everyone knows what a violin sounds like, that a large room rings and a small one does not, that a choir singing in tune is different from a choir nearly in tune. The physics does not have to argue that these effects exist. It has to explain them, and the explanations are unusually satisfying because the prediction can be checked by listening.

0.1 Audience and prerequisites

The intended reader is a university student of any major, taking this as a science course. The mathematics required is algebra, trigonometry, and logarithms: if the phrase "solve for L" and the sine function are familiar, that is enough. Calculus is not required and is not used in the main argument. Where a result is genuinely a calculus result: the wave equation from Newton's second law, Fourier coefficients as integrals, the Bessel functions that describe a drumhead: the derivation is placed in a clearly marked dropdown, stated as optional, and never depended on afterward. A reader who has had calculus will find those dropdowns are where the book stops taking their word for it; a reader who has not can skip every one and lose nothing.

No musical training is assumed. Staff notation appears only where it is the clearest picture available, and it is explained where it is first used. Readers who want a fuller treatment of the musical side will find the companion volumes *Understanding Music Theory* and *Open Music Theory* in the same series; this book cross-references them rather than duplicating them.

Instructors will find the chapters sized for assignment by numbered section. A fifteen-week pacing guide is in the Teaching Guide.

0.2 Organization

The book has six parts.

Part I, Sound and Vibration establishes the physical vocabulary: simple harmonic motion, the wave that carries a vibration through air, superposition and interference, and the standing waves and resonances that let an object have a pitch at all.

Part II, Spectrum and Timbre is the hinge of the book. Fourier's result, that any periodic sound is a sum of harmonics, is what connects the shape of an instrument to the character of its sound, and almost every later chapter uses it.

Part III, Hearing and Perception turns from the sound to the listener: the mechanics of the ear, why loudness is not intensity, why pitch is not frequency, and where consonance comes from.

Part IV, Scales and Tuning takes up the oldest quantitative problem in music. Small whole-number ratios sound good and do not fit into an octave; every tuning system in history is a different answer to that arithmetic.

Part V, Musical Instruments works through the families, strings, winds, percussion, and the voice, asking of each the same three questions: what vibrates, what sets it going, and what couples it to the air.

Part VI, Sound in Space and in Circuits follows the sound after it leaves the instrument, into the room that shapes it and the microphone that captures it.

0.3 How the chapters are built

Every chapter opens with learning objectives and a short introduction, then proceeds through numbered reading sections (§*N*.1, §*N*.2, …) kept roughly equal in length so an assignment can be stated as a range of sections. Worked examples are embedded in the reading, not quarantined at the end. Each chapter closes with a summary, a set of conceptual questions, and a problem set. Every problem opens with a tag, *Straightforward*, *Moderate*, or *Challenging*, marking how much it asks beyond a direct application of the chapter's formulas, useful when assigning a subset, or when deciding which one to try first. Solutions to every problem are included in the complete edition and omitted from the student edition.

A handful of pedagogical asides appear where they earn their keep: notes for historical context, tips for problem-solving strategy, warnings for common misconceptions, and margin remarks for a unit or a term that would otherwise interrupt the sentence. Dropdowns hold the optional derivations described above. None of these is a substitute for the main argument; all can be ignored on a first reading.

Audio examples run throughout the book. Each appears as a figure showing the waveform and spectrum of the clip, a sentence describing what is heard, and a link to hear it. They are synthesized rather than recorded, which is a deliberate limitation: a synthesized clip can isolate exactly one variable, the same pitch and the same level, differing only in which harmonics are present, in a way no recording of a real instrument can. Because every example carries its picture and its description, they work equally in print and for readers who cannot hear them. Where the real thing matters, the text says so and sends the reader to an instrument.

Interactive simulations from PhET and OpenLyceum are embedded wherever a moving picture teaches more than a static one. In the print and Word editions the same figures appear as screenshots, each caption linking to the running simulation.

0.4 The laboratory and the project

Two appendices exist because two of the course's aims cannot be met by reading. Laboratory Exercises is a set of measurements using an oscilloscope, a function generator, and audio-analysis software, or the browser simulations that stand in for all three where the hardware is missing, enough to take the book's claims about spectra, air columns, and beats and check them. The Final Project is guidance for measuring something of one's own and explaining it to somebody who does not know any physics, which is harder than it sounds and is the skill most likely to outlast the course.

0.5 A note on what is left out

This is a physics book about music, not a music book with physics in it. It has little to say about why a particular piece is beautiful, and it should not be read as claiming that the physics explains the music. What the physics explains is the *medium*: the constraints every instrument maker, tuner, singer, and hall designer has had to work within, and the reasons those constraints are what they are. Musicians solved most of these problems by ear, centuries before anyone could write down why the solutions worked. Part of the pleasure of the subject is discovering how often they were exactly right.

Part I

Sound and Vibration

Chapter 1

Sound, Music, and Simple Harmonic Motion

Learning Objectives

By the end of this chapter, you should be able to:

- Describe sound as a longitudinal pressure disturbance in a medium, and explain why sound cannot travel through a vacuum while light can.
- Distinguish the physical properties of a sound wave (frequency, amplitude, spectrum, duration) from the perceptual attributes they mostly control (pitch, loudness, timbre, length), and give an example where the correspondence fails.
- State the condition a restoring force must satisfy to produce simple harmonic motion, and explain why so many different vibrating objects satisfy it approximately.
- Write the displacement of a simple harmonic oscillator as a sinusoid, and identify its amplitude, frequency, period, and phase from a graph or an equation.
- Calculate the natural frequency of a mass on a spring and of a simple pendulum, and explain why the frequency of a mass on a spring does not depend on how hard it is struck.
- Relate the energy of a vibrating system to the square of its amplitude, and trace the exchange between kinetic and potential energy over one cycle.
- Explain what damping does to a free vibration, and connect the rate of decay to how long a note sounds after the player stops driving it.
- Estimate the frequency, wavelength, and period of a musical tone from its pitch name, and place it on the range of human hearing.

Introduction

Strike a tuning fork and hold it up. Nothing looks as though it is happening. Touch the tines and you feel a buzz, and the sound stops; let go and it starts again. Put the base of the fork against a table and the room fills with the note. Take the fork into a vacuum chamber and pump the air out, and the tines go on vibrating in plain sight while the sound dies away to nothing.

Those four observations contain most of what this chapter is about. Something has to **vibrate**. The vibration has to be passed to the **air**: a fork alone is quiet and a fork on a table is loud. Without air, or water, or wood, or some other material, there is no sound at all, however vigorously the source moves. And the sound has a definite **pitch**, the same one every time that fork is struck, whether it is struck gently or hard.

This book is about the physics that connects those facts, and it begins at the beginning: with what sound is, and with the particular kind of motion that musical sources almost always turn out to be doing. That motion is **simple harmonic motion**. Everything later, from the string and air column to the drumhead, vocal folds, microphone diaphragm, and basilar membrane inside your ear, either does it or is built from things that do.
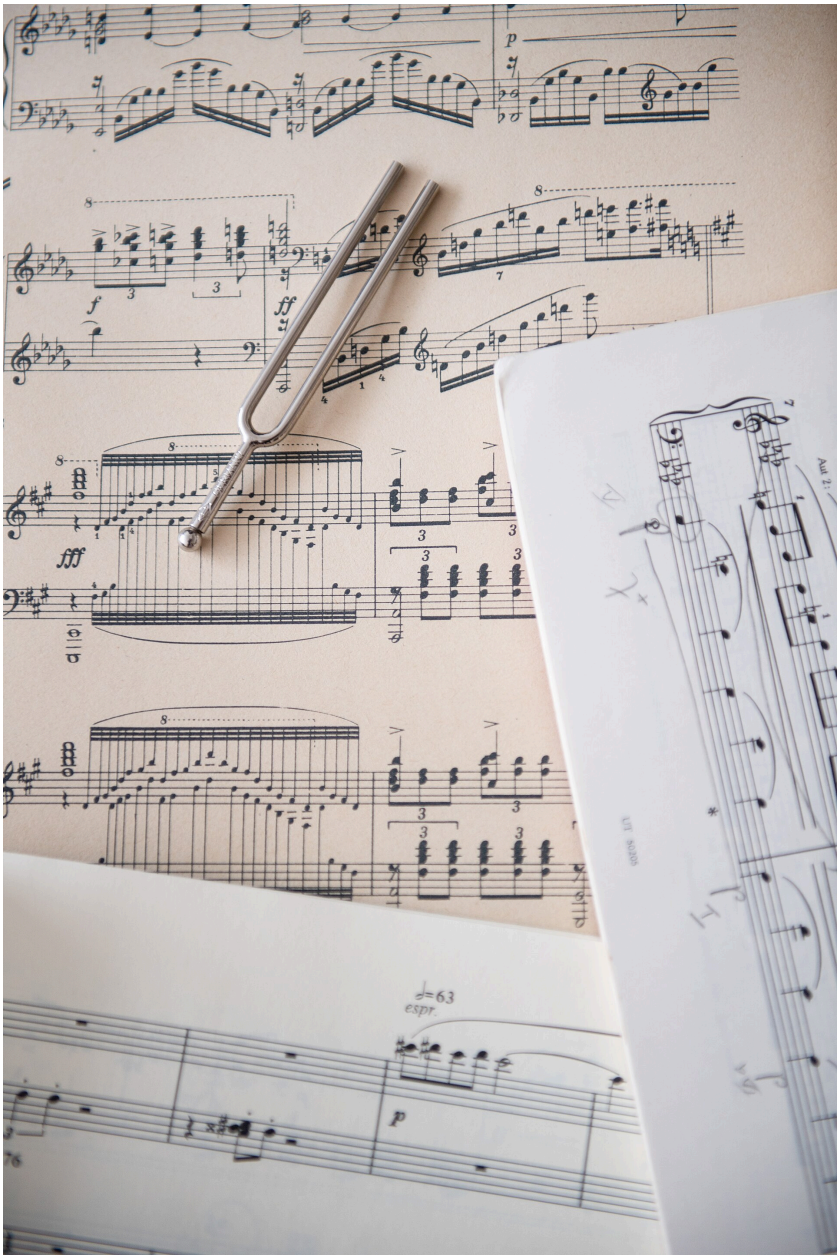


Figure 1.1: A tuning fork makes the central idea visible: a small vibrating source can have a definite musical frequency. Image: Lukasz Kobus / European Commission, CC BY 4.0.

> A vacuum chamber demonstration is easy to find on video, and worth two minutes: a ringing alarm clock or a small buzzer fades to silence as the pump runs, and comes back as air is let in.

1.1 What Sound Is

A Disturbance, Not a Substance

Sound is not a substance that travels from a source to your ear. Nothing material makes that trip. What travels is a **disturbance**: a pattern of squeezing and stretching that passes through the air, leaving the air itself essentially where it was.

The standard picture is a line of people standing shoulder to shoulder. Shove the person at one end, and they bump into the next, who bumps into the next, and a shove travels down the line. Everyone ends up roughly where they started; only the shove moves along the row. Sound is the same, with molecules of air instead of people and with collisions instead of bumping.

This matters because it explains the vacuum chamber. A disturbance needs something to disturb. Take away the air and there is nothing to pass the shove along, and no matter how hard the source vibrates, nothing arrives. Light behaves differently, and that difference is worth holding onto: light reaches us across the emptiness between stars, and sound does not cross a meter of vacuum.

> **Note**
> The nineteenth century assumed light must also need a medium, and named it the luminiferous ether. The experiments that failed to find it are the subject of relativity rather than acoustics, but the instinct behind the search is exactly the one this section is training: a wave was understood to be a disturbance *of something*.

Longitudinal Waves and Pressure

The shove down the row of people has a direction, and it is the same direction the disturbance travels. That makes a sound wave **longitudinal**: the air moves back and forth along the direction of travel, not across it.

Waves on a string are the other kind. There, the string moves up and down while the wave moves sideways, and the motion of the medium is perpendicular to the motion of the wave. Those are **transverse** waves. Both kinds matter in this book: a guitar string carries transverse waves, while the air around it carries longitudinal ones. The handover between them at the bridge and soundboard is one of the central problems of instrument design (Chapter 10).

Because the air bunches up in some places and thins out in others, sound can be described either as a **displacement** of the air or as a **pressure** variation. The two descriptions are equivalent but they are not the same picture, and confusing them is the most common early mistake in acoustics.

Where the molecules are packed more tightly than usual, the pressure is above atmospheric: a **compression**. Where they are spread thinner, the pressure is below atmospheric: a **rarefaction**. The pressure changes involved are startlingly small. Atmospheric pressure is about 101,000 Pa; the quietest sound a healthy young ear can detect is a pressure variation of about 2×10^{-5} Pa, some five thousand million times smaller. A very loud sound, a rock concert at the front, is perhaps 20 Pa, still only two ten-thousandths of the pressure it rides on. Sound is a whisper of a ripple on an ocean of static pressure.

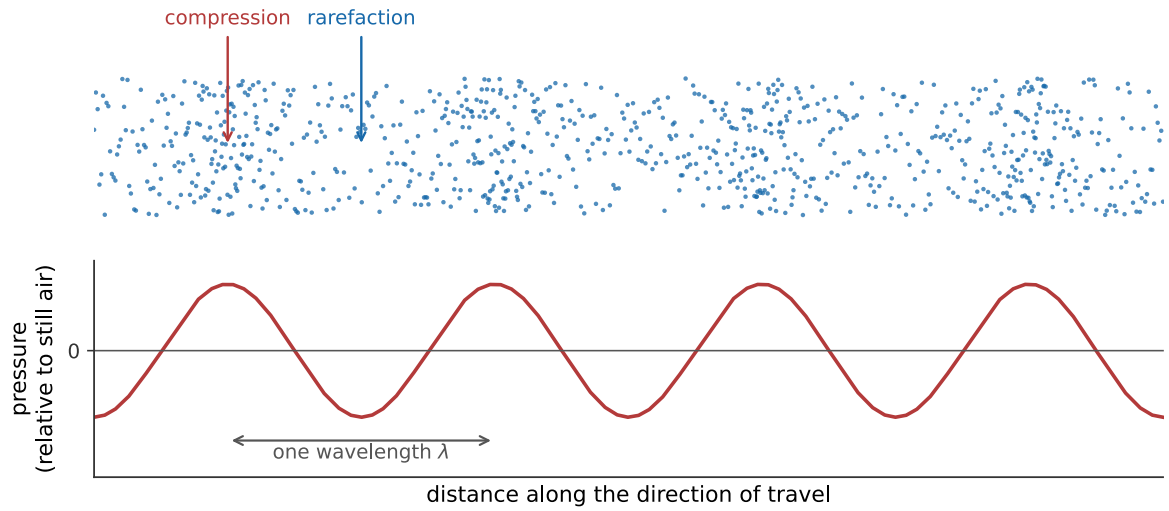


Figure 1.2: A sound wave drawn twice. Above: air molecules, bunched where the wave has pushed them together and thinned where it has pulled them apart. Below: the pressure that bunching produces. A compression is a pressure maximum and a rarefaction is a pressure minimum. Note that the pressure peaks are a quarter of a wavelength away from where the *displacement* of the air is greatest, where the air has moved furthest, it has not yet piled up.

Self-check

A loudspeaker cone moves out, then back, then out again. At the instant the cone has moved as far *out* as it will go, is the air just in front of it at maximum pressure, minimum pressure, or atmospheric? (Think about whether the cone is still compressing the air at that moment, or has stopped.)

What a Microphone and an Eardrum Measure

Both a microphone and an eardrum are pressure detectors. Each is a thin membrane with air on one side, and each is pushed in when the pressure outside rises and pushed out when it falls. What the membrane does is trace, over time, the pressure at one point in space.

That is worth dwelling on, because it means everything the rest of this book discusses is somehow encoded in a single wiggling number: pitch, loudness, timbre, the difference between a violin and a clarinet, and the sound of a concert hall: the pressure at your eardrum as a function of time. Two eardrums give two such numbers. There is nothing else. Whatever your auditory system knows about the world of sound, it works it out from that.

The graph of that pressure against time is the **waveform**, and it is the object an oscilloscope displays and an audio file stores. Chapter 5 will show that a waveform contains more structure than it looks like it does, and Chapter 6 will show how the ear takes it apart.

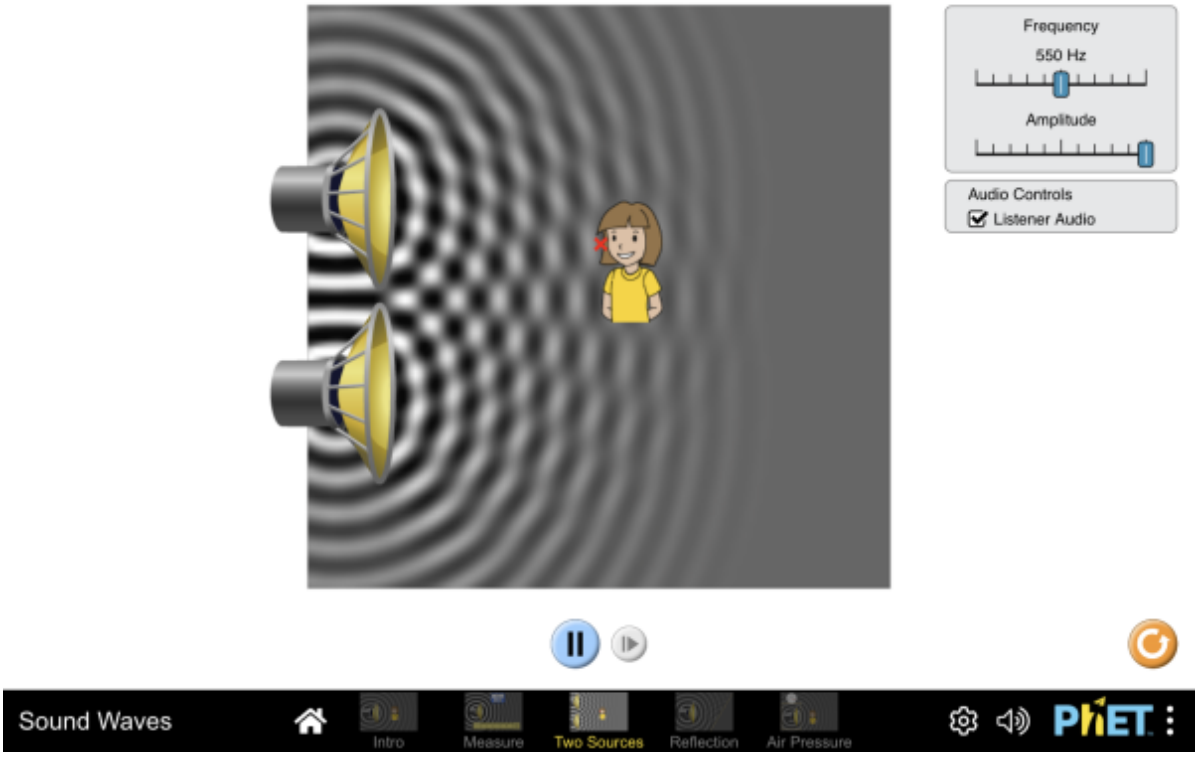


Figure 1.3: A loudspeaker driving air, with the compressions and rarefactions drawn directly. Switch between the particle view and the pressure view and check that a compression really does sit where the pressure graph peaks; then change the frequency and watch the spacing of the bands change while their speed does not. Interactive simulation: Sound Waves



Figure 1.4: The Royal Institution pumps the air from a bell jar while a bell continues to vibrate. The fading sound isolates the role of the material medium: motion at the source is not enough if there is no air to carry the pressure disturbance. Watch on YouTube: Why Sound Can't Travel in a Vacuum



Figure 1.5: Music begins with a physical performer and instrument, then becomes a signal at the ear. The guitarist is Django Reinhardt, who lost the use of the third and fourth fingers of his left hand in a caravan fire at eighteen. He rebuilt his technique around the two fingers that still worked, fretting melodies with index and middle and reaching the damaged pair across only the top strings for chords. The constraint became a style: the fast single-note lines and compact voicings of gypsy jazz. Physics fixes what a vibrating string can do, but it leaves the player remarkable freedom in how to get there. William P. Gottlieb, public domain.

1.2 From Sound to Music: Four Attributes

Pitch, Loudness, Timbre, Duration

Ask a musician to describe a note and you will get four things: how high or low it is, how loud, what it sounds like, and how long it lasts. Those are **pitch**, **loudness**, **timbre**, and **duration**, and they are the perceptual vocabulary this book has to connect to physics.

The connections are, to a first approximation:

Perceptual attribute	Physical property mostly responsible
Pitch	Frequency, how many times per second the pattern repeats
Loudness	Amplitude, how large the pressure variation is
Timbre	Spectrum and envelope, which frequencies are present, and how they change
Duration	Duration

The phrase "mostly responsible" is doing real work in that table, and the four chapters of Part III are largely about the exceptions. Loudness depends on frequency as well as amplitude: a 40 Hz tone and a 1000 Hz tone of identical amplitude are not equally loud, and are not close (Chapter 7). Pitch depends slightly on loudness. Most strikingly, a sound can have a definite pitch at a frequency that is not present in it at all (Chapter 8).

For now the first column is a useful map and the second column is where the physics is. One entry is exact: duration is duration.

Physical Cause and Perceptual Effect

The difference between the two columns needs care, because the language of music runs them together constantly and the physics does not.

Frequency is a property of the sound. Pitch is a property of your experience of it. A frequency can be measured with a counter, by anyone, with no listener present. A pitch cannot: it is a judgment, made by an auditory system, and it can be influenced by things the frequency counter knows nothing about. The same is true of amplitude and loudness, and of spectrum and timbre.

This is not pedantry. It is the reason Chapter 9 has something to explain: if pitch were simply frequency, there would be no such thing as a tuning system, only arithmetic. And it is the reason Chapter 15 can compress an audio file to a tenth of its size: the parts of the physical signal that the perceptual system cannot use can be thrown away.

> **Warning**
> "Pitch" and "frequency" are used interchangeably in ordinary musical speech, and mostly no harm is done. In this book they are kept apart. When the text says *frequency*, it means a number of cycles per second; when it says *pitch*, it means what a listener hears.

Tone and Noise

One more distinction, and then the physics. Some sounds repeat and some do not.

A sound whose waveform repeats, the same shape, over and over, many times a second, is **periodic**, and periodic sounds have a clear pitch. A sung vowel, a bowed string, a blown pipe: hold any of them steady and the waveform settles into a pattern that repeats. The number of repeats per second is the **frequency**, measured in hertz (Hz), and one cycle takes the **period** $T = 1/f$.

A sound whose waveform does not repeat is **noise**, and noise has no definite pitch. A cymbal crash, a handclap, a consonant, the hiss of air escaping: the waveform wanders and never comes back to where it was.

Music uses both, and the boundary is not sharp. A snare drum is mostly noise, a flute is mostly tone, and the breathy attack of a flute note is noise on the front of a tone. Chapter 5 makes this quantitative; Chapter 12 is largely about instruments that sit between the two.

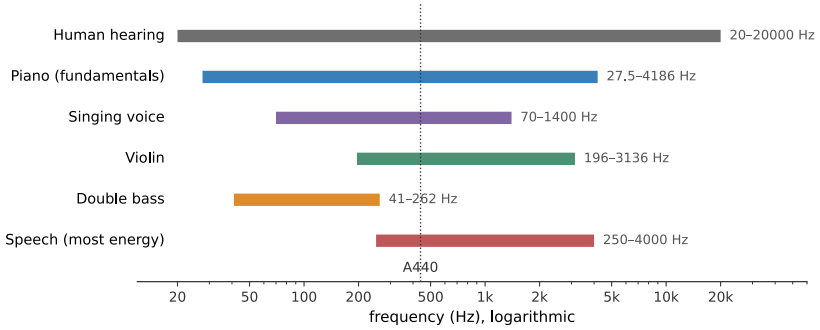


Figure 1.6: Where music sits inside the range of hearing. The frequency axis is logarithmic, because that is how pitch works: each doubling of frequency is one octave, and the octaves are therefore equally spaced along this axis. A piano's lowest note is 27.5 Hz and its highest is 4186 Hz, seven octaves, and less than half the span of human hearing, most of which is above the highest note any orchestral instrument plays.

Two things in that figure deserve comment. First, the audible range is about 20 Hz to 20,000 Hz: a factor of a thousand, or very nearly ten octaves. Second, almost all musical *fundamentals* live in the bottom half of it. The upper octaves are not empty, but what fills them is the harmonics that give instruments their character (Chapter 5). A recording that discards everything above 5 kHz therefore sounds muffled rather than merely missing a few high notes.

1.3 Simple Harmonic Motion

The Linear Restoring Force

Something has to vibrate. The question is why so many different somethings vibrate in the same way.

Take any object sitting at a stable equilibrium: a mass hanging on a spring, a pendulum at the bottom of its swing, a guitar string pulled taut, a diving board, a wine glass rim. "Stable" means that if you displace it slightly, something pushes it back. Call the displacement x and the restoring force F. The force is zero at $x = 0$, by the definition of equilibrium, and it points back toward $x = 0$ on either side.

Now the key step. For *small* displacements, the restoring force of essentially any stable system is proportional to the displacement:

$$F = -kx. \tag{1.1}$$

The constant k is the **stiffness**, measured in newtons per meter, and the minus sign says the force opposes the displacement. This is **Hooke's law**, and the reason it applies so widely is not that nature is fond of springs. It is that any smooth restoring force, whatever its true form, looks like a straight line if you zoom in far enough on the point where it crosses zero, and small vibrations never leave that neighborhood.

Why any stable system obeys Hooke's law for small displacements

This is the argument that justifies the whole chapter, and it needs one line of calculus.

Let the potential energy of the system be $U(x)$, with a stable equilibrium at $x = 0$. Expand it as a Taylor series:

$$U(x) = U(0) + U'(0)\,x + \tfrac{1}{2}U''(0)\,x^2 + \tfrac{1}{6}U'''(0)\,x^3 + \cdots \tag{1.2}$$

$U(0)$ is a constant and can be set to zero; it does not affect the force. $U'(0) = 0$, because the force $F = -\mathrm{d}U/\mathrm{d}x$ vanishes at equilibrium, that is what equilibrium means. So the first surviving term is the quadratic one, and

$$F = -\frac{\mathrm{d}U}{\mathrm{d}x} = -U''(0)\,x - \tfrac{1}{2}U'''(0)\,x^2 - \cdots \tag{1.3}$$

For small enough x, the x^2 term and everything after it are negligible beside the x term, and what remains is $F = -kx$ with $k = U''(0)$. Stability requires $U''(0) > 0$, a minimum, not a maximum, so k is positive.

The generality is the point: nothing was assumed about the system beyond having a smooth potential with a minimum. A mass on a spring, a molecule in a crystal, and a drumhead all vibrate sinusoidally at small amplitude for that reason, and "small amplitude" is the caveat on nearly every result in this book. Where it fails, a cymbal driven hard (Chapter 12), a loudspeaker pushed past its limits (Chapter 15), the consequences are audible and specific.

The Sinusoid and Its Four Numbers

A system obeying $F = -kx$ moves in a particular way, and only that way. Displaced and released, it traces out a sine curve in time:

$$x(t) = A\sin(2\pi f t + \phi). \tag{1.4}$$

This is **simple harmonic motion**, and the equation has exactly four numbers in it. Three describe the motion and the fourth describes when you started the clock.

- **Amplitude** A is the maximum displacement, measured from equilibrium. It is half the peak-to-trough distance, and confusing the two is worth a factor of two in any calculation that follows.
- **Frequency** f is the number of complete cycles per second, in hertz. Its reciprocal is the **period** $T = 1/f$, the time one cycle takes.
- **Phase** ϕ says where in the cycle the motion is at $t = 0$. On its own it means nothing, shifting your clock changes it, but the phase *difference* between two sinusoids is physical and important, and it is the whole subject of Chapter 3.

You will also meet the **angular frequency** $\omega = 2\pi f$, which is measured in radians per second and exists to keep 2πs out of equations. In terms of it, $x(t) = A\sin(\omega t + \phi)$.

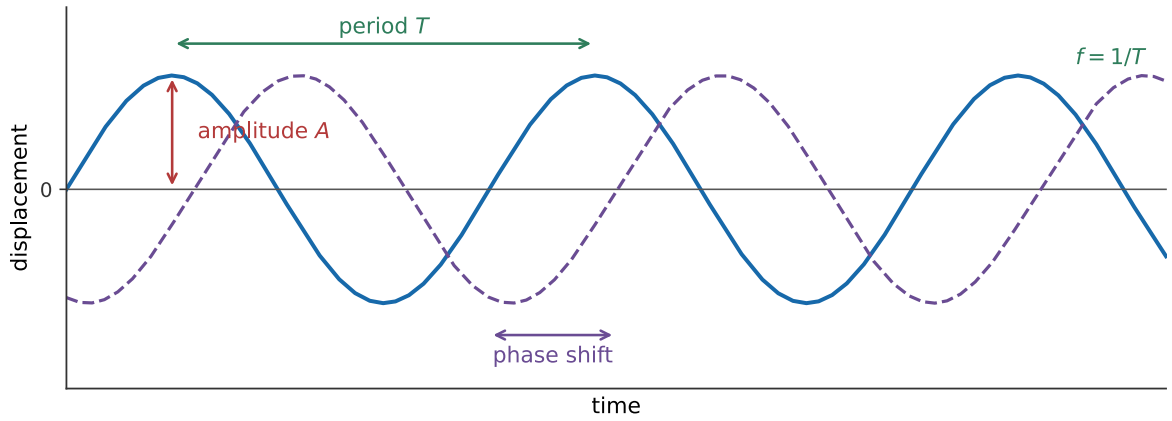


Figure 1.7: The anatomy of a sinusoid. The amplitude is measured from the equilibrium line to the peak, not from trough to peak. The period is the time for one complete cycle, and the frequency is its reciprocal. The dashed curve has the same amplitude and the same frequency and differs only in phase, it is the same motion, started at a different moment.

Both conventions are common: f in hertz is what a tuner displays and what a musician names; ω in radians per second is what appears in the algebra. Convert with $\omega = 2\pi f$ and check units when in doubt.

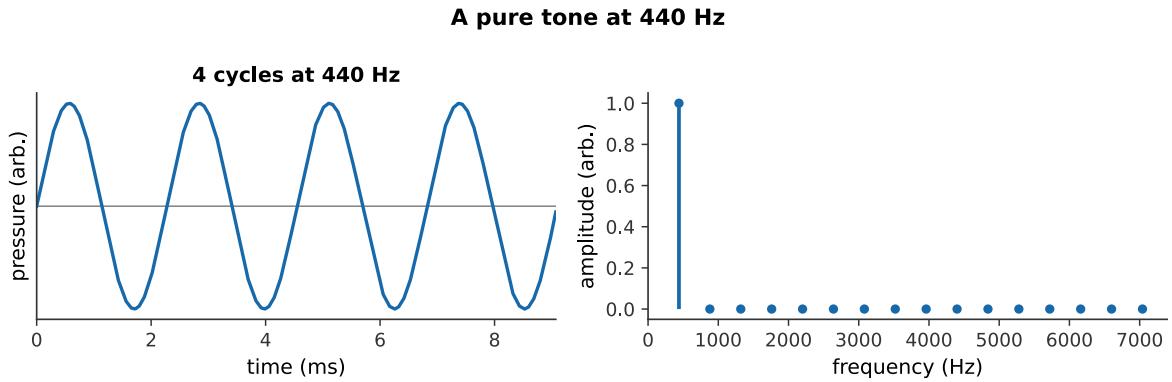


Figure 1.8: A pure tone at 440 Hz: a single sinusoid, and nothing else. Almost nothing in music sounds like this, and that is the point, every musical instrument in this book produces something more complicated, and Chapter 5 explains what it is made of. *What you hear:*A steady, plain, slightly thin tone at concert A. It sounds a little like a hearing test, which is exactly what a pure tone is. Listen: A440

Mass on a Spring

The simplest system that obeys Hooke's law exactly is a mass m on a spring of stiffness k, and its frequency is

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}. \tag{1.5}$$

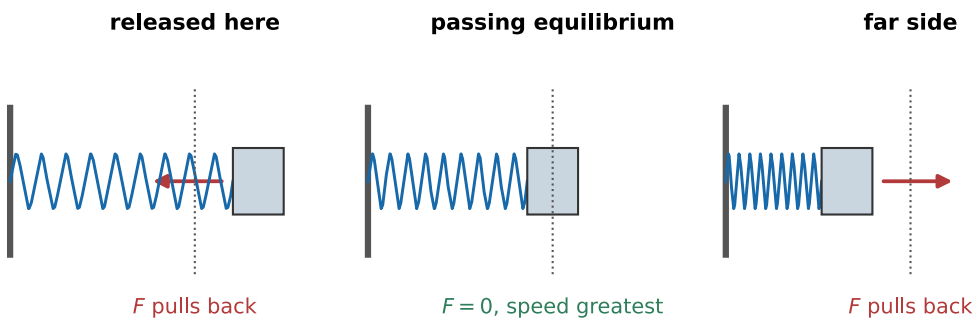


Figure 1.9: One cycle of simple harmonic motion, in three snapshots. The restoring force is largest where the displacement is largest and the mass is momentarily at rest; it is zero as the mass passes through equilibrium, where the speed is greatest. Force and displacement are always in opposite directions, which is the entire content of $F = -kx$.

Two features of that formula are worth more than the formula itself.

Stiffer means faster; heavier means slower. Tighten the spring and the frequency rises; hang more mass on it and the frequency falls. Every pitch-changing mechanism in every instrument in this book is a version of one of those two levers. A guitar string is tuned by changing its tension, stiffness, and a bass string is made low by winding metal around it, mass. The vocal folds go up in pitch when their tension rises.

The frequency does not depend on the amplitude. Pull the mass twice as far and release it, and it takes exactly as long to come back. This is a peculiar property of the *linear* restoring force, and it is not a general fact about vibrating things. But it is the property that makes music possible at all: a piano string struck gently and struck hard sounds the same note. If frequency depended on amplitude, every instrument would go sharp or flat as it got louder, and no fixed-pitch instrument could exist.

Worked example: tuning by mass and by tension

A mass of 0.20 kg on a spring oscillates at 2.5 Hz.

(a) What is the spring's stiffness?

Rearranging $f = (1/2\pi)\sqrt{k/m}$ gives $k = m(2\pi f)^2$, so

$$k = (0.20\text{ kg})\,[2\pi(2.5\text{ Hz})]^2 = (0.20\text{ kg})(15.7\text{ s}^{-1})^2 = 49\text{ N/m}. \tag{1.6}$$

(b) What mass would oscillate at exactly half that frequency on the same spring?

Frequency goes as $1/\sqrt{m}$, so halving f requires multiplying m by four: 0.80 kg.

(c) A guitar's low E string is tuned to 82.4 Hz. Its maker wants a string one octave lower, 41.2 Hz, of the same length and the same tension. By what factor must its mass per unit length increase?

The same $1/\sqrt{m}$ scaling applies to a string (Chapter 3 derives it), so halving the frequency requires four times the mass per unit length. Winding wire around a thin core is how this is done in practice; making the core four times as massive by making it thicker would also make it far too stiff to bend over the bridge, which is the subject of §10.1.

The Pendulum, and the Small-Angle Approximation

A pendulum is the other classic example, and it is instructive precisely because it obeys Hooke's law only approximately. A bob of mass m on a string of length L, displaced by an angle θ, feels a restoring force $F = -mg\sin\theta$ along its arc. That is not proportional to θ, but for small angles, $\sin\theta \approx \theta$, and it becomes so. The frequency is then

$$f = \frac{1}{2\pi}\sqrt{\frac{g}{L}}, \tag{1.7}$$

which, strikingly, does not contain the mass at all.

The approximation is good to better than 1% out to about 14°, and it fails slowly rather than suddenly: at 30° the real period is about 1.7% longer than the formula predicts, and a pendulum swinging through 90° is nearly 18% slow. This is the general behavior of a real vibrating system driven beyond its linear range; keep it in mind when Chapter 12 reaches instruments that are never in it.

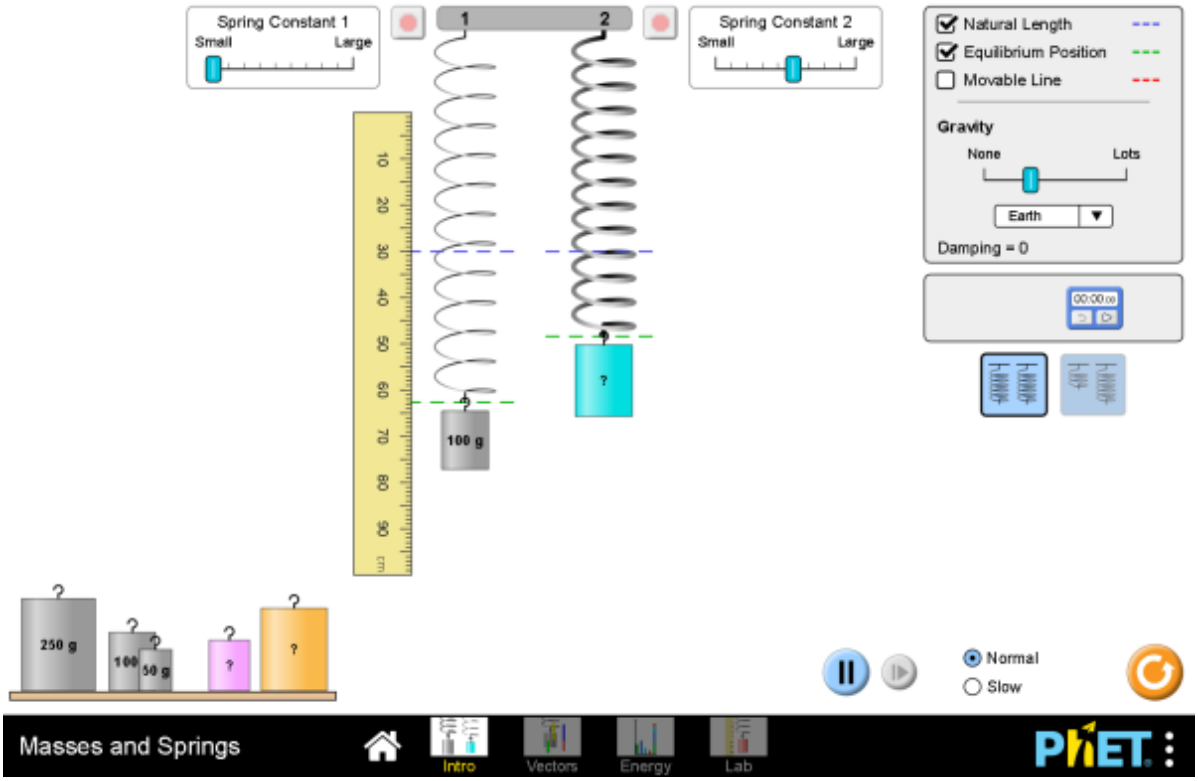


Figure 1.10: Hang masses on springs of different stiffness and time the oscillations. Two experiments are worth doing deliberately: change the amplitude while keeping the mass and spring fixed, and confirm that the period does not move; then quadruple the mass and confirm that the period exactly doubles. Interactive simulation: Masses and Springs



Figure 1.11: Brusspup's pendulum array turns small differences in period into traveling patterns, beating, and repeated returns to alignment. Each bob is an independent oscillator; the apparent wave comes from their steadily changing relative phases. Watch on YouTube: Amazing Pendulum Wave Effect!

1.4 Energy, Amplitude, and Damping

Energy Goes as Amplitude Squared

A vibrating system carries energy, and it carries it in two forms that trade back and forth. At the extremes of the motion the mass is momentarily at rest and all the energy is stored in the stretched spring; as it passes through equilibrium the spring is relaxed and all the energy is kinetic. In between, it is some of each. For a system obeying $F = -kx$, the total is

$$E = \tfrac{1}{2}kA^2, \tag{1.8}$$

and the fact to carry forward is the **square**. Doubling the amplitude of a vibration quadruples its energy. A sound wave of twice the pressure amplitude carries four times the energy, and delivers four times the power to your eardrum.

That squaring is why Chapter 7 needs logarithms. The range of sound energies the ear handles is not merely large, it is the *square* of an already large range of pressures, about 10^{12} to one, from the quietest audible sound to the threshold of pain. No linear scale is usable across that span, and the decibel exists to tame it.

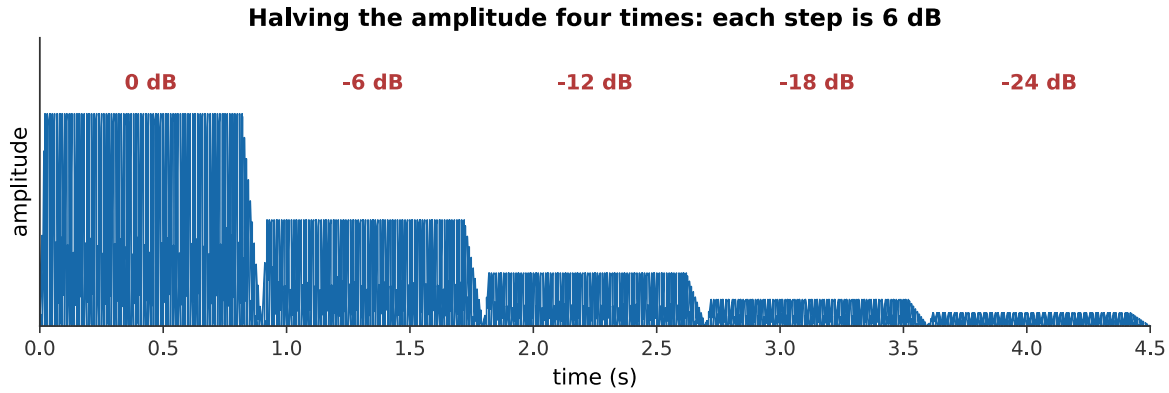


Figure 1.12: The same 440 Hz tone at five amplitudes, each half the one before. Each halving of amplitude is a factor of four in energy, and a drop of 6 dB. Notice how *un*dramatic four steps sound: a sixteen-fold reduction in amplitude, a 256-fold reduction in energy, is not remotely "a sixteenth as loud". Making that observation precise is the business of Chapter 7. *What you hear:*One tone, five times, each quieter than the last by a clearly audible but not dramatic step. The last is about a sixteenth the amplitude of the first, and still perfectly audible. Listen: Amplitude Steps

Damped Vibration and the Decay of a Note

An undamped oscillator would ring forever. Real ones do not, because energy leaks away: into friction inside the material, into the mountings, and, usefully, since this is the point of a musical instrument, into the surrounding air as sound.

The result is that the amplitude falls off over time, and for light damping it falls off *exponentially*:

$$A(t) = A_0 e^{-t/\tau}. \tag{1.9}$$

The **decay time** τ is the time for the amplitude to fall to $1/e$, about 37%, of where it started. A large τ means a long ring; a small τ means a quick thud.

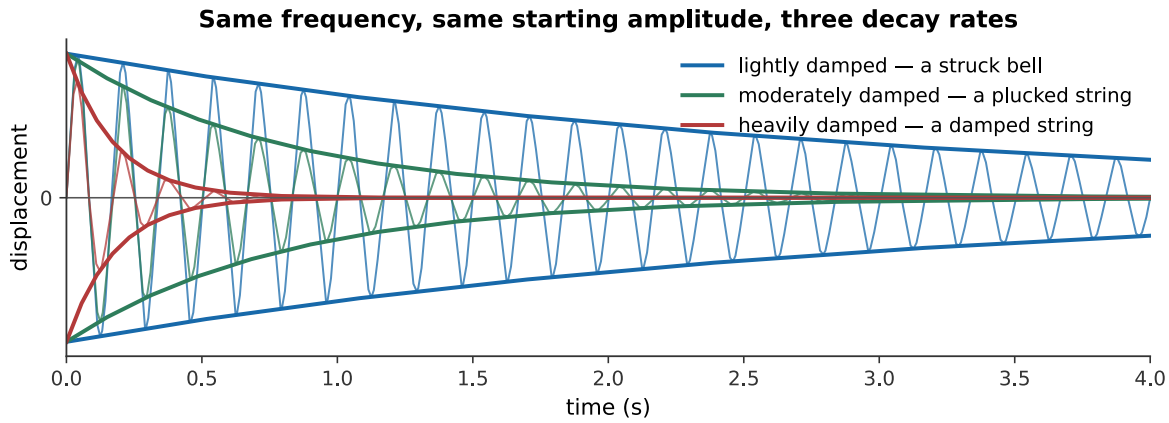


Figure 1.13: Three decay rates at the same frequency and the same starting amplitude. The frequency is barely affected by damping this light, the zero crossings stay where they were, but the length of the note is completely transformed. An instrument's sustain is therefore a design parameter independent of its pitch.

Every instrument makes a choice here, and the choice is a trade-off that Chapter 4 makes precise. Energy radiated as sound is energy lost from the vibration, so **a loud instrument is a short-lived one**. A guitar string coupled tightly to a good soundboard is loud and dies quickly; the same string on a solid plank is quiet and rings for a long time. A pianist who wants a longer note lifts the dampers; a pianist who wants a shorter one puts them back.

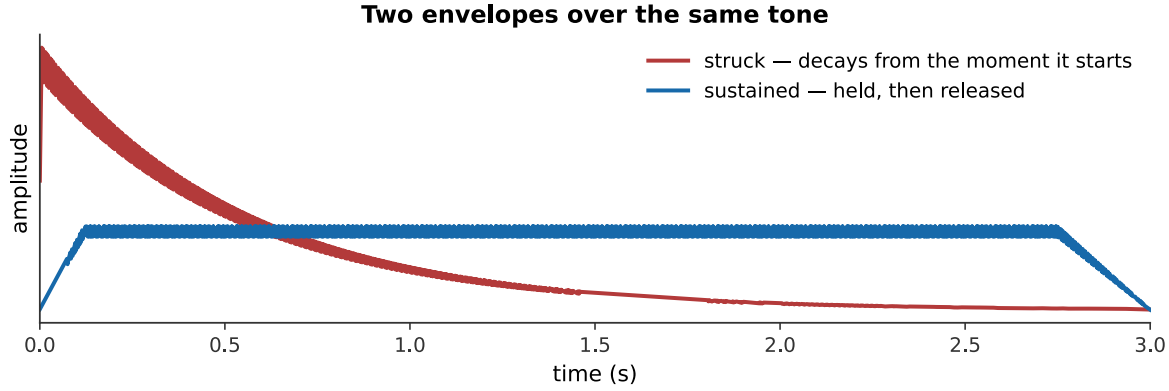


Figure 1.14: The same pitch, the same harmonics, the same loudness, and two envelopes. Nothing differs between these but how the amplitude changes with time, and it is enough to make the first sound struck and the second sound bowed or blown. Chapter 5 shows that this is not a minor effect: strip the attack off a recorded note and listeners struggle to name the instrument. *What you hear:*Two notes of the same pitch and the same tone color. The first begins abruptly and fades away like a plucked string; the second grows in, holds steady, and stops. Listen: Struck, Sustained

Why Anything Vibrates Sinusoidally At All

Why does simple harmonic motion deserve a chapter to itself, given that almost nothing in music is a mass on a spring?

The answer has three parts, and each is developed later in the book.

First, everything stable is a spring for small displacements. That is the Taylor-series argument above. Air in a pipe, a stretched membrane, a steel bar, the fluid in your cochlea: displace any of them slightly and the restoring force is linear in the displacement, so the motion is sinusoidal.

Second, complicated objects are collections of simple ones. A guitar string does not have one natural frequency; it has an unlimited number, and each of them behaves exactly like an independent mass on a spring. These are its **normal modes**, and Chapter 4 shows that any motion of any such system, however complicated, is a sum of its modes vibrating simultaneously.

Third, any periodic sound is a sum of sinusoids. This is Fourier's theorem, the subject of Chapter 5, and it is the reason a chapter about the simplest possible vibration is the right place to start a book about the most complicated ones. A clarinet's waveform looks nothing like a sine curve. It is nevertheless a sum of sine curves, and once you know which ones, you know why a clarinet sounds like a clarinet.

Simple harmonic motion is not an approximation the book will later outgrow. It is the alphabet.

1.5 Summary

- **Sound is a longitudinal disturbance in a medium**, not a substance in transit. The medium moves back and forth along the direction of travel and stays where it was; only the disturbance goes anywhere. No medium, no sound, light crosses a vacuum and sound does not.
- **A sound wave can be described as displacement or as pressure**, and the two are a quarter of a wavelength out of step. Compressions are pressure maxima and rarefactions are pressure minima. Both microphones and eardrums measure pressure, so the whole of what a listener has to work with is pressure at a point as a function of time.
- **Musical sounds are described by four perceptual attributes**, pitch, loudness, timbre, duration, which correspond roughly to frequency, amplitude, spectrum-and-envelope, and duration. Only the last correspondence is exact, and the exceptions are the subject of Part III.
- **Periodic waveforms have pitch; non-periodic ones are noise.** Frequency f and period T are reciprocals, $T = 1/f$. Human hearing spans roughly 20 Hz to 20 kHz; musical fundamentals occupy only the lower half of it.
- **Any stable system obeys $F = -kx$ for small displacements**, because any smooth potential looks quadratic near its minimum. That is how so many unlike objects come to vibrate in the same way, and why "small amplitude" qualifies nearly every result in this book.
- **Simple harmonic motion is a sinusoid**, $x(t) = A\sin(2\pi ft + \phi)$, described by amplitude, frequency, and phase. For a mass on a spring $f = (1/2\pi)\sqrt{k/m}$: stiffer is faster, heavier is slower, and **the frequency does not depend on the amplitude**, without which no fixed-pitch instrument could exist.
- **Energy goes as the square of the amplitude**, $E = \frac{1}{2}kA^2$. The enormous dynamic range this produces is why loudness is measured logarithmically (Chapter 7).
- **Damping makes amplitude decay exponentially**, $A(t) = A_0 e^{-t/\tau}$. Radiating sound is a loss of energy from the vibration, so loudness and sustain trade against each other: a design decision every instrument maker has to make.

1.6 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity. The set mixes multiple choice, true/false, drag-the-words, fill-in-the-blank, and mark-the-words formats, and opens on the anatomy of a sinusoid.

1. **Multiple choice, from a figure.** The activity reproduces the anatomy-of-a-sinusoid figure from earlier in this chapter, which draws two sinusoids of equal amplitude and equal period. Which quantity distinguishes them: phase, amplitude, frequency, or period?

2. **True or false.** Doubling the amplitude of a vibrating string doubles the energy stored in the vibration.

3. **Drag the words.** Complete: in a _____ the air is bunched and the pressure is _____ atmospheric; in a _____ the air is spread out and the pressure is _____ atmospheric; an eardrum and a microphone both respond to _____. (Terms: compression, above, rarefaction, below, pressure, displacement, equal to.)

4. **Fill in the blanks.** A steady tone at 250 Hz has a period of _____ ms, because frequency and period are _____.

5. **Mark the words.** Identify the three numbers that specify a simple harmonic motion: "A sinusoid is fixed by its amplitude, its frequency, and its phase; loudness, pitch, and timbre are the perceptions those give rise to."

1.7 Conceptual Questions

1. An alarm clock ringing inside a bell jar goes silent as the air is pumped out, but the hammer can still be seen striking the bell. Explain what has stopped and what has not.

2. A tuning fork held in the air is quiet; the same fork with its base pressed against a table is loud, and stops ringing sooner. Account for both observations with one explanation.

3. A sound wave is sometimes drawn as a wavy line. Explain what is actually wavy, given that the air does not move up and down at all, and say what quantity the vertical axis of such a drawing represents.

4. At the moment a small parcel of air has been displaced furthest from its rest position, is the pressure there a maximum, a minimum, or atmospheric? Explain your answer without using calculus.

5. A piano string struck hard and struck gently produces the same pitch. Which property of simple harmonic motion guarantees this, and what would music be like if it did not hold?

6. Two guitar strings have the same length and the same tension, but one is twice as massive per unit length. Which sounds the lower note, and by roughly what interval?

7. A pendulum clock keeps good time when its swing is small, but runs slow when the swing is large. Explain which assumption of §1.3 has failed, and in which direction the error goes.

8. Explain, in terms of where the vibrational energy goes, why an electric guitar unplugged is both quieter and longer-sustaining than an acoustic guitar.

1.8 Problems

Exercise 1.1
(Straightforward) A tuning fork vibrates at 512 Hz. (a) What is its period? (b) How many complete cycles does it make in the 2.4 s it takes to fade to inaudibility?

Solution to Exercise 1.1
(a) The period is the reciprocal of the frequency:

$$T = \frac{1}{f} = \frac{1}{512\ \text{Hz}} = 1.95 \times 10^{-3}\ \text{s} = 1.95\ \text{ms}. \tag{1.10}$$

(b) The number of cycles is the frequency times the time:

$$N = ft = (512\ \text{s}^{-1})(2.4\ \text{s}) = 1.2 \times 10^{3}\ \text{cycles}. \tag{1.11}$$

Therefore, the period is 1.95 ms and the fork completes about 1200 cycles.

Exercise 1.2
(Straightforward) The A above middle C is tuned to 440 Hz. (a) What is the frequency of the A one octave above it, and one octave below? (b) What is the frequency four octaves above 440 Hz, and is it audible?

Solution to Exercise 1.2
(a) An octave is a factor of two in frequency. One octave above: $2 \times 440 = 880$ Hz. One octave below: $440/2 = 220$ Hz.
(b) Four octaves is a factor of $2^4 = 16$:

$$f = 16 \times 440\ \text{Hz} = 7040\ \text{Hz}. \tag{1.12}$$

Therefore, the note is 7040 Hz, which is well inside the audible range of roughly 20 Hz to 20 kHz, though it is above the highest note on a piano (4186 Hz).

Exercise 1.3
(Straightforward) A mass of 0.35 kg hangs from a spring of stiffness $k = 84$ N/m. (a) At what frequency does it oscillate? (b) What mass would be needed to halve that frequency?

Solution to Exercise 1.3
(a) Using $f = (1/2\pi)\sqrt{k/m}$:

$$f = \frac{1}{2\pi}\sqrt{\frac{84\ \text{N/m}}{0.35\ \text{kg}}} = \frac{1}{2\pi}\sqrt{240\ \text{s}^{-2}} = \frac{15.5}{6.283} = 2.47\ \text{Hz}. \tag{1.13}$$

(b) Since $f \propto 1/\sqrt{m}$, halving the frequency requires quadrupling the mass:

$$m = 4 \times 0.35\ \text{kg} = 1.4\ \text{kg}. \tag{1.14}$$

Therefore, the frequency is 2.47 Hz, and 1.4 kg would halve it.

Exercise 1.4

(Straightforward) A simple harmonic oscillator has amplitude $A = 3.0$ mm and frequency $f = 250$ Hz. (a) Write its displacement as a function of time, taking $x = 0$ at $t = 0$ and moving in the positive direction. (b) What is its maximum speed?

Solution to Exercise 1.4

(a) Starting at $x = 0$ and moving positive means a sine with zero phase:

$$x(t) = (3.0 \times 10^{-3}\ \text{m})\sin\!\left[2\pi(250\ \text{Hz})\,t\right] = (3.0\ \text{mm})\sin\!\left[(1571\ \text{s}^{-1})\,t\right]. \tag{1.15}$$

(b) For simple harmonic motion the maximum speed is $v_{\max} = \omega A = 2\pi f A$:

$$v_{\max} = 2\pi(250\ \text{Hz})(3.0 \times 10^{-3}\ \text{m}) = 4.7\ \text{m/s}. \tag{1.16}$$

Therefore, the displacement is $x(t) = (3.0\ \text{mm})\sin[(1571\ \text{s}^{-1})t]$ and the maximum speed is 4.7 m/s, which is reached as it passes through equilibrium, where the restoring force is zero.

Exercise 1.5

(Straightforward) The threshold of hearing corresponds to a pressure amplitude of about 2.0×10^{-5} Pa, and the threshold of pain to about 20 Pa. (a) What is the ratio of these pressures? (b) What is the ratio of the *energies* they carry?

Solution to Exercise 1.5

(a) The pressure ratio is

$$\frac{20\ \text{Pa}}{2.0 \times 10^{-5}\ \text{Pa}} = 1.0 \times 10^{6}. \tag{1.17}$$

(b) Energy goes as the square of amplitude, so

$$\left(1.0 \times 10^{6}\right)^{2} = 1.0 \times 10^{12}. \tag{1.18}$$

Therefore, the ear handles a millionfold range of pressures, which is a million-million-fold range of energies. A linear scale covering this span is unusable, which is the reason for the decibel (Chapter 7).

Exercise 1.6

(Moderate) A guitar string's vibration decays with $\tau = 1.8$ s. (a) What fraction of its initial amplitude remains after 3.0 s? (b) After how long has the amplitude fallen to 10% of its initial value? (c) What fraction of the initial *energy* remains at that moment?

Solution to Exercise 1.6

(a) Using $A(t) = A_0 e^{-t/\tau}$:

$$\frac{A}{A_0} = e^{-3.0/1.8} = e^{-1.667} = 0.189. \tag{1.19}$$

(b) Setting $A/A_0 = 0.10$ and solving:

$$e^{-t/\tau} = 0.10 \;\Rightarrow\; \frac{t}{\tau} = \ln 10 = 2.303 \;\Rightarrow\; t = 2.303 \times 1.8\ \text{s} = 4.1\ \text{s}. \tag{1.20}$$

(c) Energy goes as amplitude squared, so at 10% amplitude the energy is $(0.10)^2 = 1.0\%$ of its initial value.

Therefore, about 19% of the amplitude remains after 3.0 s, 10% remains after 4.1 s, and at that point only one part in a hundred of the energy is left, yet the note is still clearly audible, which is another demonstration of the ear's range.

Exercise 1.7

(Moderate) Two identical springs, each of stiffness k, support the same mass m, first one at a time, then both side by side in parallel. By what factor does the oscillation frequency change when the second spring is added?

Solution to Exercise 1.7

Two springs in parallel both stretch by the same displacement x, and each contributes a restoring force kx, so the total force is $2kx$. The effective stiffness is therefore $k_{\text{eff}} = 2k$.

Since $f \propto \sqrt{k}$,

$$\frac{f_2}{f_1} = \sqrt{\frac{2k}{k}} = \sqrt{2} = 1.41. \tag{1.21}$$

Therefore, the frequency rises by a factor of $\sqrt{2}$, about 41%, which, as Chapter 9 will point out, is almost exactly a musical tritone, the interval of six equal-tempered semitones.

Exercise 1.8

(Straightforward) A pendulum is to be built with a period of exactly 1.00 s. (a) How long must it be? Take $g = 9.81$ m/s^2. (b) The same pendulum is taken to the Moon, where $g = 1.62$ m/s^2. What is its period there?

Solution to Exercise 1.8

(a) From $f = (1/2\pi)\sqrt{g/L}$ and $T = 1/f$, the period is $T = 2\pi\sqrt{L/g}$. Rearranging:

$$L = g\left(\frac{T}{2\pi}\right)^2 = (9.81\ \text{m/s}^2)\left(\frac{1.00\ \text{s}}{6.283}\right)^2 = 0.248\ \text{m}. \tag{1.22}$$

(b) The period goes as $1/\sqrt{g}$:

$$T_{\text{Moon}} = T_{\text{Earth}}\sqrt{\frac{g_{\text{Earth}}}{g_{\text{Moon}}}} = (1.00\ \text{s})\sqrt{\frac{9.81}{1.62}} = 2.46\ \text{s}. \tag{1.23}$$

Therefore, the pendulum is 24.8 cm long and swings with a period of 2.46 s on the Moon. Note that the mass of the bob never entered either calculation.

Exercise 1.9

(Straightforward) A loudspeaker cone moves in simple harmonic motion at 80 Hz with an amplitude of 1.5 mm. (a) What is the maximum speed of the cone? (b) What is its maximum acceleration, in units of g?

Solution to Exercise 1.9

First, $\omega = 2\pi f = 2\pi(80\text{ Hz}) = 503\text{ s}^{-1}$.

(a) The maximum speed is

$$v_{\max} = \omega A = (503\text{ s}^{-1})(1.5 \times 10^{-3}\text{ m}) = 0.75\text{ m/s}. \tag{1.24}$$

(b) For simple harmonic motion the maximum acceleration is $a_{\max} = \omega^2 A$:

$$a_{\max} = (503\text{ s}^{-1})^2(1.5 \times 10^{-3}\text{ m}) = 380\text{ m/s}^2, \tag{1.25}$$

which is $380/9.81 = 39$ times g.

Therefore, the cone reaches 0.75 m/s and $39g$. Accelerations of this order are routine in loudspeakers, and they are the reason cones are made as light and as stiff as materials allow (Chapter 15).

Exercise 1.10

(Moderate) A note played on a violin has a waveform that repeats every 2.27 ms. (a) What is its frequency? (b) Which note is it, given that A440 lies 440 Hz and each semitone is a factor of $2^{1/12}$? (Count semitones from A440.)

Solution to Exercise 1.10

(a) The frequency is the reciprocal of the period:

$$f = \frac{1}{2.27 \times 10^{-3}\text{ s}} = 441\text{ Hz}. \tag{1.26}$$

(b) The ratio to A440 is $441/440 = 1.0023$. One semitone is $2^{1/12} = 1.0595$, so the number of semitones is

$$n = \frac{\ln 1.0023}{\ln 1.0595} = \frac{0.00227}{0.0578} = 0.039\text{ semitones}. \tag{1.27}$$

Therefore, the note is A440, sharp by about four hundredths of a semitone, four **cents**, in the units of Chapter 9. That is far too small to hear as mistuning, and is well within the precision with which a violinist places a finger.

Exercise 1.11

(Challenging) Show that for a simple harmonic oscillator the speed at displacement x is

$$v = \omega\sqrt{A^2 - x^2}, \tag{1.28}$$

and use it to find at what fraction of the amplitude the kinetic and potential energies are equal.

Solution to Exercise 1.11

Energy is conserved, and the total is the potential energy at the turning point:

$$\tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = \tfrac{1}{2}kA^2. \tag{1.29}$$

Solving for v:

$$v^2 = \frac{k}{m}\left(A^2 - x^2\right) = \omega^2\left(A^2 - x^2\right), \tag{1.30}$$

using $\omega^2 = k/m$. Taking the square root gives $v = \omega\sqrt{A^2 - x^2}$, which correctly gives $v = \omega A$ at $x = 0$ and $v = 0$ at $x = \pm A$.

The two energies are equal when each is half the total:

$$\tfrac{1}{2}kx^2 = \tfrac{1}{2}\left(\tfrac{1}{2}kA^2\right) \;\Rightarrow\; x^2 = \tfrac{1}{2}A^2 \;\Rightarrow\; x = \frac{A}{\sqrt{2}} = 0.707A. \tag{1.31}$$

Therefore, the energy is evenly split at about 71% of the amplitude, not at half the amplitude, which is the natural but wrong guess.

Exercise 1.12

(Moderate) A piano string and a wine glass are both struck and left to ring. The string's amplitude falls to half in 0.9 s; the glass's falls to half in 4.5 s. (a) Find τ for each. (b) Which radiates sound more efficiently, and how do you know? (c) Estimate how long each remains audible if a note becomes inaudible once its amplitude has fallen by a factor of 10^3.

Solution to Exercise 1.12

(a) Setting $A/A_0 = \tfrac{1}{2} = e^{-t_{1/2}/\tau}$ gives $\tau = t_{1/2}/\ln 2$:

$$\tau_{\text{string}} = \frac{0.9\ \text{s}}{0.693} = 1.3\ \text{s}, \qquad \tau_{\text{glass}} = \frac{4.5\ \text{s}}{0.693} = 6.5\ \text{s}. \tag{1.32}$$

(b) The string. A short decay time means energy is leaving the vibration quickly, and for an instrument designed to be heard, the dominant loss channel is radiation as sound. The glass holds its energy, which is why a wine glass is loud only when it is rubbed continuously, and why the string, losing its energy fast, is loud straight away.

(c) A factor of 10^3 takes $t = \tau \ln(10^3) = 6.91\tau$:

$$t_{\text{string}} = 6.91 \times 1.3\ \text{s} = 9.0\ \text{s}, \qquad t_{\text{glass}} = 6.91 \times 6.5\ \text{s} = 45\ \text{s}. \tag{1.33}$$

Therefore, the string rings for about 9 s and the glass for about 45 s. The comparison is the trade-off of §1.4 in numbers: the glass sustains five times longer because it is five times worse at the job of turning vibration into sound.

Exercise 1.13

(Moderate) A student claims that because sound is a pressure wave, a loudspeaker must push a "packet of air" from the cone to the listener's ear. (a) Estimate how long a packet of air would take to travel 5 m if it moved at the maximum cone speed of 0.75 m/s found in Exercise 1.9. (b) Sound actually covers 5 m in about 15 ms. Use the comparison to explain what is wrong with the student's picture.

Solution to Exercise 1.13

(a) At a steady 0.75 m/s:

$$t = \frac{5\ \text{m}}{0.75\ \text{m/s}} = 6.7\ \text{s}. \tag{1.34}$$

(b) The actual travel time is about 15 ms, some 440 times shorter. Worse, the cone's motion is *oscillatory*, it moves out and back, so no parcel of air makes net progress at all; each one jiggles about a fixed position with an amplitude of a fraction of a millimeter.

Therefore, the student's picture is wrong in both magnitude and kind. What travels from the loudspeaker to the ear is not air but a *disturbance* in the air, and its speed is set

by how quickly molecular collisions pass the squeeze along: a property of the medium, not of the source (Chapter 2).

Exercise 1.14

(Moderate) A mass–spring system is set oscillating with amplitude A. A second, identical system is set oscillating with amplitude $3A$. (a) Compare their frequencies. (b) Compare their maximum speeds. (c) Compare their energies. (d) Which comparison explains why a piano stays in tune as the pianist plays louder?

Solution to Exercise 1.14

(a) Identical, since $f = (1/2\pi)\sqrt{k/m}$ contains no amplitude.

(b) $v_{\max} = \omega A$, so the second is three times faster.

(c) $E = \tfrac{1}{2}kA^2$, so the second has $3^2 = 9$ times the energy.

(d) Part (a). The independence of frequency from amplitude is exactly what guarantees that a string struck harder sounds the same note, only louder. Were it otherwise, the instrument would go out of tune with every change of dynamic, and fixed-pitch instruments could not exist.

Chapter 2

Wave Motion and the Speed of Sound

Learning Objectives

By the end of this chapter, you should be able to:

- Distinguish transverse from longitudinal waves and identify which each of a string, an air column, and a drumhead supports.
- Apply the relation $v = f\lambda$ to find any one of wave speed, frequency, or wavelength given the other two, and explain which of the three a musician changes when playing a different note.
- Explain why the speed of a wave is set by the medium rather than by the source, and predict what happens to frequency and to wavelength when a sound crosses from one medium into another.
- Calculate the speed of sound in air as a function of temperature, and estimate the pitch error a wind instrument suffers when it is played cold.
- Compare the speed of sound in gases, liquids, and solids using the stiffness-to-density ratio, and explain why sound travels faster in helium than in air.
- Define sound intensity, apply the inverse-square law to a point source in the open air, and explain why it fails indoors.
- Define acoustic impedance and explain, in terms of impedance mismatch, why very little sound crosses from a vibrating string directly into air.
- Describe the Doppler effect for a moving source and for a moving listener, calculate the shifted frequency, and identify where it is heard in music.

Introduction

Chapter 1 left a vibrating object pushing on the air next to it. This chapter follows what happens next.

The answer is not obvious. The sound arrives at your ear far faster than any air could get there, and the air that arrives is not the air that left. A trumpet player thirty meters away blows a note, and a fifteenth of a second later you hear it, but the breath that made it is still in the trumpet. Something crossed the room, and it was not a substance.

What crossed the room was a **wave**, and this chapter is about how fast it goes, what it looks like on the way, and what happens to it as it spreads. Three results will be used constantly for the rest of the book. The first is the relation $v = f\lambda$, which connects the frequency a musician

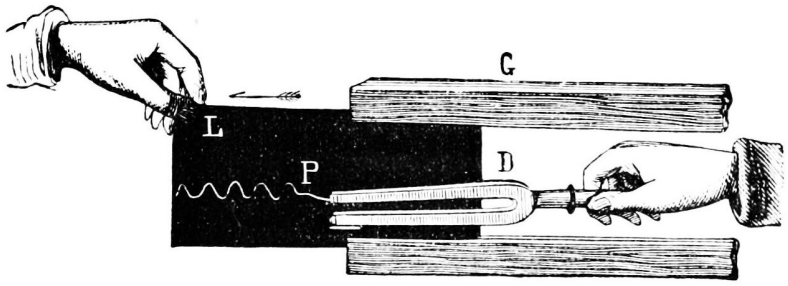


Figure 2.1: An early illustration of a tuning fork producing a visible sound vibration. Public-domain illustration from *Popular Science Monthly* (1878).

cares about to the wavelength the physics cares about. The second is that the speed v is a property of the *medium*, not of the source, not of the frequency, not of how loud the sound is, so an orchestra stays together and a cold clarinet plays flat. The third is that sound spreads out, and spreading has a cost, which is the beginning of the story of why instruments need bodies and bells at all.

2.1 From a Vibration to a Wave

How a Disturbance Propagates

Return to the row of people standing shoulder to shoulder. Shove the first one and a shove travels down the line. Two questions are worth asking about it.

What decides how fast the shove travels? Two things. If the people are **stiff**, they resist being compressed and push back hard: the shove passes quickly. If they are **heavy**, hard to get moving, it passes slowly. Nothing else matters: not how hard you shoved, not how often you shove, not how far down the line you look.

What decides how far the shove goes? In an idealized line, nothing; it goes forever. In reality some energy is lost to friction at each step, and the shove fades. Both effects have acoustic counterparts, and the first, speed set by stiffness and inertia, is the one that runs this chapter.

The same reasoning applies to air. Air is stiff because compressing it raises its pressure, and it has inertia because it has mass. The ratio of those two properties fixes the speed of sound, and §2.3 makes the statement quantitative.

Transverse and Longitudinal

Two kinds of wave appear throughout this book, and every instrument uses both.

The distinction is not cosmetic. It determines what a wave can do: only transverse waves can be polarized, so light can and sound cannot. It also determines what has to happen at a boundary between the two. That handover, from the transverse motion of a string to the longitudinal motion of air, turns out to be the hardest engineering problem in the design of a string instrument, and §2.4 explains why.

> A wave on the surface of water is neither, quite: a floating cork traces a small circle rather than a straight line. Surface waves are a mixture, and a poor analogy for sound even though everyone reaches for them first.

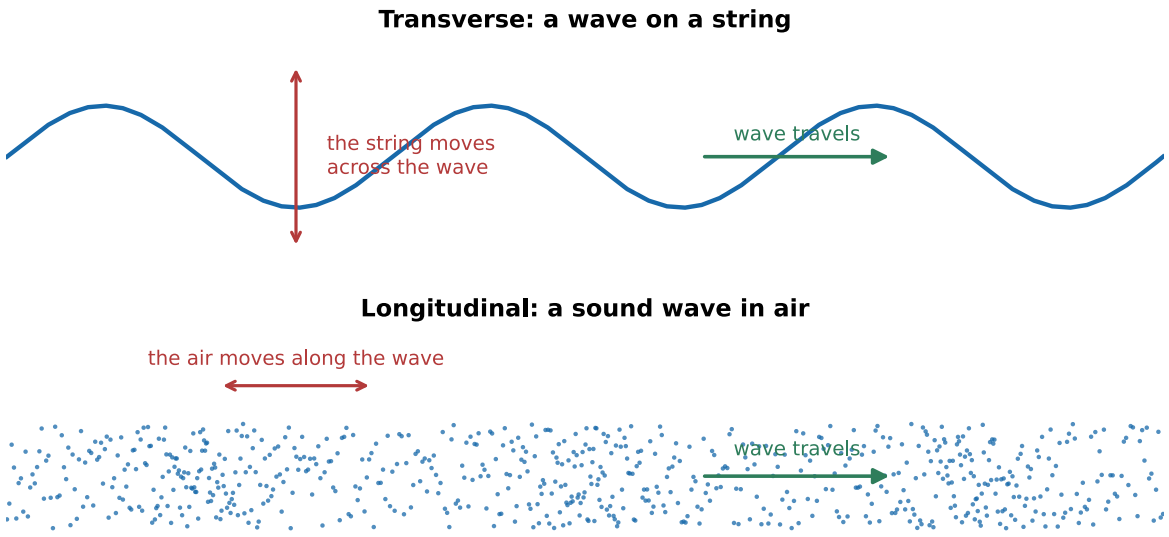


Figure 2.2: The two kinds of wave. In a **transverse** wave the medium moves across the direction of travel; in a **longitudinal** wave it moves along it. A guitar string carries transverse waves, the air around it carries longitudinal ones, and the instrument's job is to get energy from the first into the second.

The Wave Pulse and the Periodic Wave

Shove the row of people once and a single **pulse** travels down it. Shove them rhythmically and a **periodic wave** travels instead: a repeating pattern, moving at the same speed the single pulse would have moved.

A handclap is a pulse. A sung note is a periodic wave. Musically the periodic case is the interesting one, but the pulse is the more fundamental: a periodic wave is what you get by sending pulses one after another, and its speed is the pulse speed. **The speed does not know or care about the frequency.** This is worth stating plainly because it is the single most useful fact in the chapter, and because the exceptions to it, media in which speed *does* depend on frequency, called dispersive media, are rare in air and consequential when they occur. A piano string is mildly dispersive, and Chapter 10 shows that this small fact is why pianos are tuned the way they are.

2.2 Wavelength, Frequency, and Wave Speed

The Relation $v = f\lambda$

A periodic wave has a repeating pattern in space as well as in time. The **wavelength** λ is the distance between one crest and the next; the **period** T is the time between one crest and the next passing a fixed point.

In one period, the wave advances exactly one wavelength. So its speed is

$$v = \frac{\lambda}{T} = f\lambda. \tag{2.1}$$

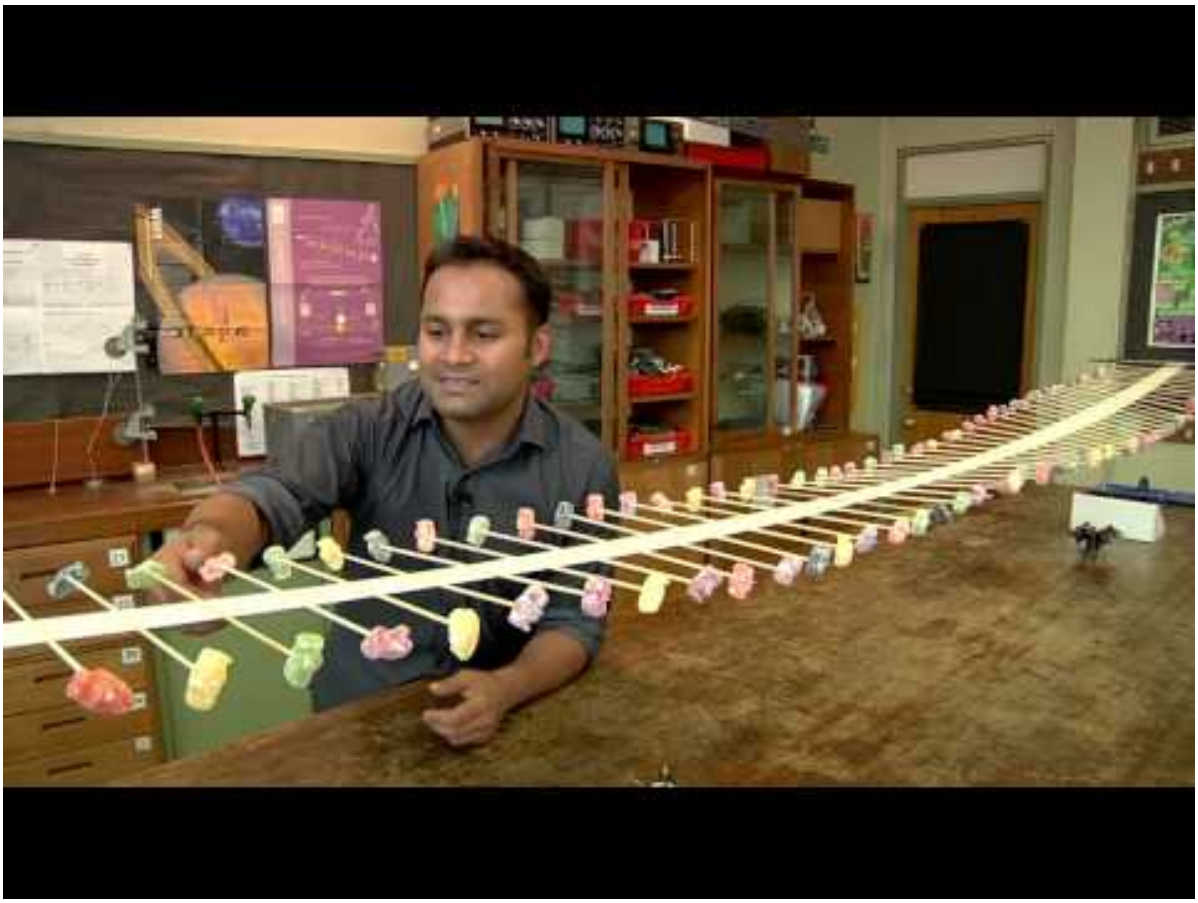


Figure 2.3: The National STEM Center's tabletop wave machine makes propagation visible: each element moves locally while the disturbance and its energy travel along the machine. Reflections and changes of medium can be seen without following an individual piece of matter down the line. Watch on YouTube: Wave Machine Demonstration

That is the whole derivation, and the equation carries an enormous amount of this book.

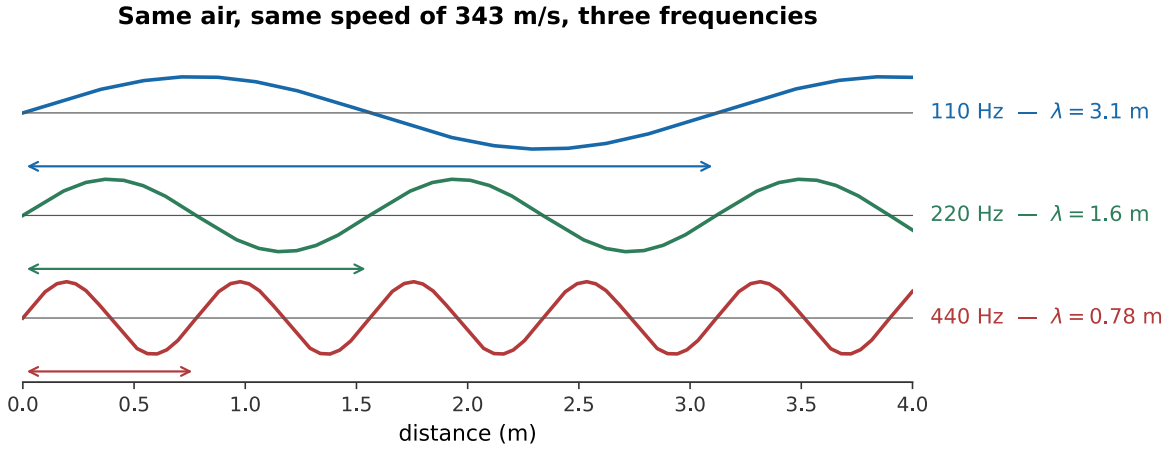


Figure 2.4: Three notes in the same air. The speed is the same for all three, 343 m/s, so a higher frequency must mean a shorter wavelength, in exactly inverse proportion. Doubling the frequency, which is going up an octave, halves the wavelength.

Notice the *sizes*. Musical wavelengths in air are comparable to the dimensions of rooms and of people: a low A on a bass is about eight meters long, middle C about 1.3 m, and the top note of a piccolo about 8 cm. Sound therefore bends around corners and light does not; bass notes are hard to control in a small room (Chapter 14); and a loudspeaker that is small compared with a wavelength radiates in all directions while one that is large does not.

Worked example: wavelengths across the piano

The speed of sound in air at 20 °C is 343 m/s. Find the wavelengths of the lowest and highest notes on a piano, 27.5 Hz and 4186 Hz.

$$\lambda_{\text{low}} = \frac{v}{f} = \frac{343\ \text{m/s}}{27.5\ \text{Hz}} = 12.5\ \text{m}, \qquad \lambda_{\text{high}} = \frac{343\ \text{m/s}}{4186\ \text{Hz}} = 8.2\ \text{cm}. \tag{2.2}$$

The lowest note's wavelength is longer than most rooms; the highest note's is the width of a hand. That range, a factor of 152, is the reason no single loudspeaker driver can reproduce a piano, and the reason bass is the hardest part of a room to get right.

A Wave Crossing a Boundary

When a wave passes from one medium into another, one of its three quantities is forced to stay the same and the other two must adjust.

Frequency is conserved. The far side of the boundary is driven by the near side, wiggle for wiggle; the boundary cannot swallow cycles or invent them. Whatever the frequency was, it still is.

Speed changes, because the speed belongs to the medium.

Wavelength therefore changes too, since $\lambda = v/f$ with f fixed.

Sound entering water from air quadruples its wavelength, because water carries sound over four times as fast. A singer's note does not become a different note when it reaches a listener underwater; it is the same pitch, stretched out in space.

> **Warning**
> It is tempting to think that "the note changes" when sound enters a new medium, because the speed changes and $v = f\lambda$ has f in it. It does not. Fix on what the boundary physically does: it shakes the next medium at the rate it is being shaken, so the conserved quantity is obvious.

The Wave Equation

Everything above can be compressed into one statement, which is the standard starting point of a more mathematical treatment. Any disturbance $y(x,t)$ that travels at a fixed speed v without changing shape satisfies

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}. \tag{2.3}$$

This is the **wave equation**. Nothing in the rest of this book requires it, and it will not be used again in the main text; it is stated so the derivation can answer the question "why is the speed what it is?"

Where the wave equation comes from, and why $v = \sqrt{T/\mu}$

Take a stretched string of tension T and mass per unit length μ, displaced by a small amount $y(x,t)$, and apply Newton's second law to a short segment between x and $x + \mathrm{d}x$.

The tension pulls on each end of the segment along the string. Because the string is curved, the two pulls do not quite cancel: the vertical components are $T\sin\theta$ at each end, and for small slopes $\sin\theta \approx \tan\theta = \partial y/\partial x$. The net upward force is therefore the *difference* of the slopes at the two ends,

$$\mathrm{d}F = T\left[\left.\frac{\partial y}{\partial x}\right|_{x+\mathrm{d}x} - \left.\frac{\partial y}{\partial x}\right|_{x}\right] = T\,\frac{\partial^2 y}{\partial x^2}\,\mathrm{d}x. \tag{2.4}$$

The segment's mass is $\mu\,\mathrm{d}x$ and its vertical acceleration is $\partial^2 y/\partial t^2$, so Newton's second law reads

$$\mu\,\mathrm{d}x\,\frac{\partial^2 y}{\partial t^2} = T\,\frac{\partial^2 y}{\partial x^2}\,\mathrm{d}x, \qquad \text{that is} \qquad \frac{\partial^2 y}{\partial t^2} = \frac{T}{\mu}\,\frac{\partial^2 y}{\partial x^2}. \tag{2.5}$$

Comparing with the wave equation identifies the speed:

$$v = \sqrt{\frac{T}{\mu}}. \tag{2.6}$$

Stiffness over inertia, exactly as §2.1 argued. Tension is what pulls the string back; mass per unit length is what resists being moved. This is the formula Chapter 3 uses to work out the pitch of a string, and it contains no reference to amplitude or frequency: the small-slope approximation is what removes them.

To confirm that the equation really describes traveling waves, substitute $y = f(x - vt)$ for any twice-differentiable f. The chain rule gives $\partial^2 y/\partial t^2 = v^2 f''$ and $\partial^2 y/\partial x^2 = f''$, so the equation is satisfied for *any* shape f. That is the mathematical statement that a pulse travels without changing shape, and that its speed does not depend on what shape it is.

2.3 What Sets the Speed of Sound

Stiffness and Inertia

The pattern of the dropdown above is general. For any medium,

$$v = \sqrt{\frac{\text{stiffness}}{\text{density}}}, \tag{2.7}$$

where "stiffness" is whatever resists the deformation the wave makes, and "density" is whatever has to be accelerated. For a string it is tension over mass per unit length. For a gas it is the bulk modulus over the mass density. For a solid bar it is Young's modulus over density.

Two consequences follow immediately, and both are counterintuitive on first meeting.

Denser is slower, other things being equal. Heavier things are harder to shake. Sound travels faster in helium than in air not because helium is thinner in some vague sense but because it is about seven times less dense while being similarly stiff.

Stiffer is faster, and stiffness wins. Steel is eight thousand times denser than air, and sound travels through it seventeen times *faster*, because steel is some two million times stiffer. When a medium is both denser and stiffer, the stiffness usually wins by more.

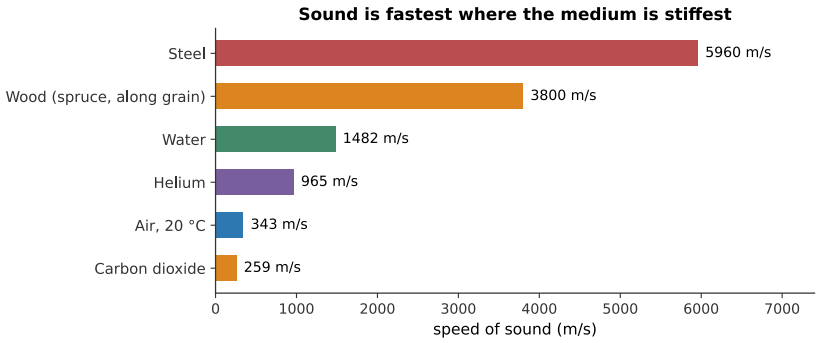


Figure 2.5: The speed of sound in six media. The ordering is not by density. Steel is the densest thing here and the fastest, so the order follows the ratio of stiffness to density. Spruce is anisotropic: sound runs along the grain about four times faster than across it, which is exactly why soundboards are cut the way they are (Chapter 10).

Sound in Air, and the Temperature Dependence

For air at ordinary temperatures, the speed of sound is very nearly

$$v = 331.3 + 0.606\,T \quad \text{m/s}, \tag{2.8}$$

with T the temperature in degrees Celsius. At 0 °C sound travels at 331 m/s; at 20 °C, at 343 m/s; at 35 °C, at 353 m/s.

The dependence is on temperature and essentially nothing else. Air pressure does not appear, which surprises people: raising the pressure makes air both stiffer and denser, in the same proportion, and the two effects cancel exactly. Humidity has a small effect, moist air is very slightly faster, because water molecules are lighter than the nitrogen they displace, but it is a fraction of a percent and rarely matters.

Temperature matters a great deal, because a wind instrument's sounding length is fixed and its pitch follows the speed of sound directly.

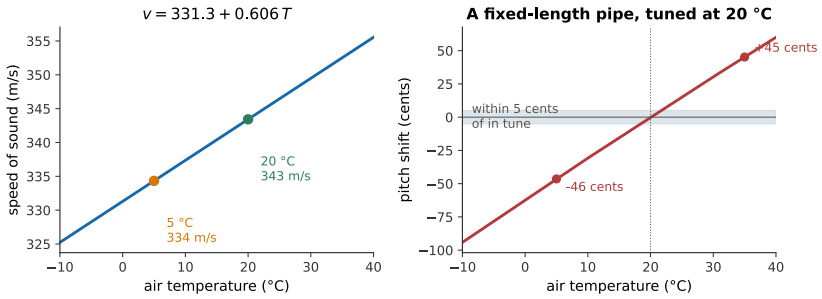


Figure 2.6: Left: the speed of sound in air against temperature. Right: what that does to a pipe of fixed length, expressed in **cents**, hundredths of a semitone, the units of Chapter 9. A pipe tuned at 20 °C and played at 5 °C is about 46 cents flat, very nearly a quarter of a semitone, and unmistakably out of tune.

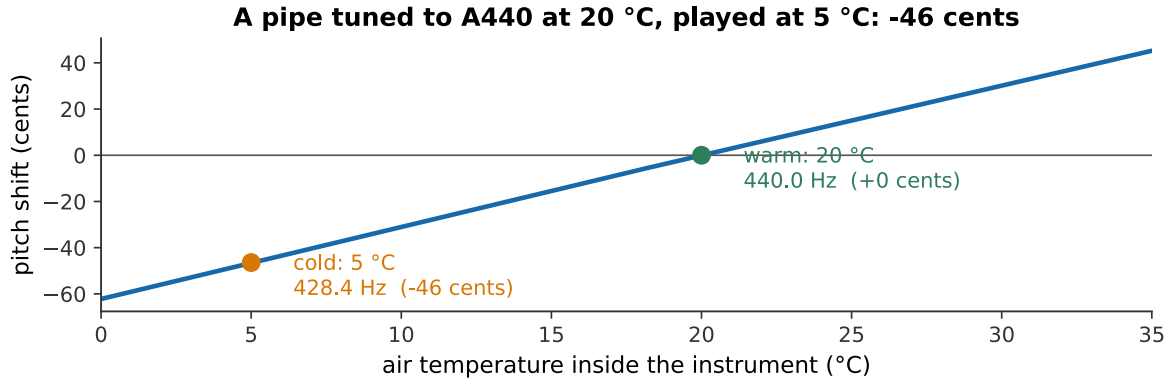


Figure 2.7: A pipe of fixed length, cold and warm. Wind players blow warm air through their instruments before a performance for this reason, and an orchestra retunes after the interval because the strings have not moved while the winds have warmed up and gone sharp. *What you hear:*The same instrument twice. The second is noticeably higher than the first, not by a semitone, but clearly enough that the two would clash if played together. Listen: At 5 °C, At 20 °C

> **Note**
> The effect works in the other direction for strings. A guitar left in a cold car goes *sharp*, because the steel strings contract and their tension rises. A temperature change therefore pushes the winds and the strings in opposite directions, and a mixed ensemble in a cold hall is harder to tune than either section alone.

Sound in Liquids and Solids

Sound travels about 1482 m/s in water, some 4.3 times its speed in air. In solids it is faster still, and in a wooden soundboard it depends sharply on direction: along the grain of spruce, about 3800 m/s; across the grain, about a quarter of that. Instrument makers have exploited this for

centuries without needing the number: a soundboard is cut so the grain runs the length of the instrument, because that is the direction along which vibrations must be distributed quickly.

Worked example: locating a hammer blow along a rail

A steel rail is struck at one end. A listener 600 m away, with one ear to the rail, hears two distinct sounds. How far apart are they?

Through the steel, at 5960 m/s:

$$t_{\text{steel}} = \frac{600\ \text{m}}{5960\ \text{m/s}} = 0.101\ \text{s}. \tag{2.9}$$

Through the air, at 343 m/s:

$$t_{\text{air}} = \frac{600\ \text{m}}{343\ \text{m/s}} = 1.75\ \text{s}. \tag{2.10}$$

The gap is $1.75 - 0.10 = 1.65$ s, easily long enough to hear as two separate events, which is the classic demonstration that the speed of sound belongs to the medium and not to the source.

2.4 Sound Spreading in Three Dimensions

Intensity and the Inverse-Square Law

So far the wave has been drawn as though it traveled along a line. In the open air it does not: it spreads out in all directions, and spreading dilutes it.

Intensity I is the power carried by a wave per unit area, in watts per square meter. A source radiating power P uniformly in all directions spreads that power over a sphere, so at distance r

$$I = \frac{P}{4\pi r^2}. \tag{2.11}$$

This is the **inverse-square law**, and it is a statement about geometry rather than about sound: double the distance and the same power is spread over four times the area, so the intensity falls to a quarter. In the units of Chapter 7 that is a drop of 6 dB per doubling of distance.

The indoor case is the more common one in music, and the difference is worth stating plainly: **inside a room, the inverse-square law fails beyond a few meters.** Someone at the back of a concert hall is not hearing a sound 40 dB weaker than the front row; they are hearing a sound perhaps 8 dB weaker, most of which has bounced off something.

Directivity: Why Instruments Do Not Radiate Evenly

The inverse-square law assumed a source radiating equally in all directions. Real instruments do no such thing.

The controlling quantity is the size of the source compared with the wavelength. A source much *smaller* than a wavelength radiates nearly uniformly, it has no way to favor one direction, because the whole of it is effectively at one point as far as the wave is concerned. A source

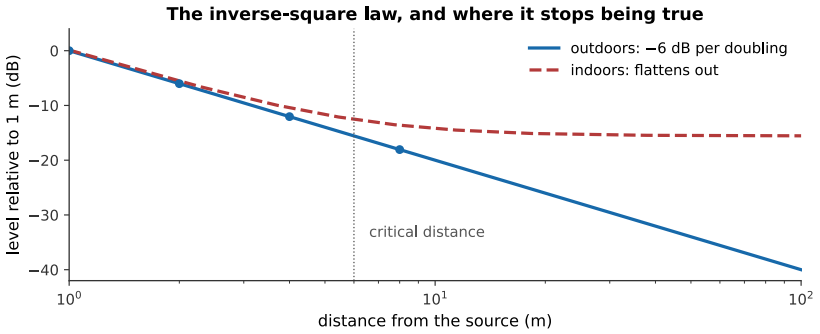


Figure 2.8: Level against distance. Outdoors, the level falls 6 dB for every doubling, a straight line on these axes. Indoors it does not: beyond the **critical distance**, the reflected sound bouncing around the room is stronger than the sound arriving directly, and moving further away stops making much difference. A lecturer at the back of a reverberant hall is audible but unintelligible for just this reason; Chapter 14 takes up the indoor case in detail.

much *larger* than a wavelength beams: the different parts of it interfere, and the sound goes mostly where they all agree.

Since instruments are fixed in size and wavelength shrinks as pitch rises, **every instrument becomes more directional at high frequencies**. A trumpet's low notes spill around the hall; its high notes go where the bell points, so a trumpet sounds quite different from the side. A cello radiates low notes in all directions and high notes in a complicated pattern of lobes. Recording engineers know this as microphone placement: moving a microphone 30 cm can change a recorded timbre more than changing the microphone does.

Acoustic Impedance and the Coupling Problem

There is one more property of a medium that matters, and it explains a puzzle raised in Chapter 1: why is a vibrating tuning fork almost silent until you touch it to a table?

Acoustic impedance measures how hard a medium is to set moving: the ratio of the pressure applied to the flow that results. A stiff, dense medium like steel has a high impedance: large pressures produce small motions. A light, springy medium like air has a low one: small pressures produce large motions.

Sound crosses a boundary efficiently only when the impedances on the two sides are **similar**. When they are very different, almost all of the energy reflects and almost none is transmitted. The mismatch between a solid and air is enormous, and the consequence is that a thin vibrating object, a string, a tine, a reed, is a terrible radiator of sound. It shakes the air, the air offers almost no resistance, and hardly any energy is handed over.

Every acoustic instrument is, in large part, a solution to this problem. A soundboard is a large light surface that presents an impedance intermediate between string and air. A brass bell is a gradual flare that converts a narrow high-impedance column into a wide low-impedance one. The middle ear, remarkably, is a mechanical lever-and-piston arrangement solving the same problem in reverse, matching air to the fluid of the inner ear (Chapter 6).

> **See Also**
> The impedance argument is used four more times in this book, each time as the key to a different instrument: String Instruments: Guitar, Violin, and Piano for the bridge and

soundboard, Wind Instruments and Air-Column Resonance for the bell, The Ear and the Physiology of Hearing for the middle ear, and Electronic and Recorded Sound for the loudspeaker.

2.5 The Doppler Effect

Moving Source, Moving Listener

If the source of a sound is moving, the wavefronts it leaves behind are not evenly spaced. Each crest is emitted from a slightly different place, so ahead of the source they are bunched closer together and behind it they are stretched further apart.

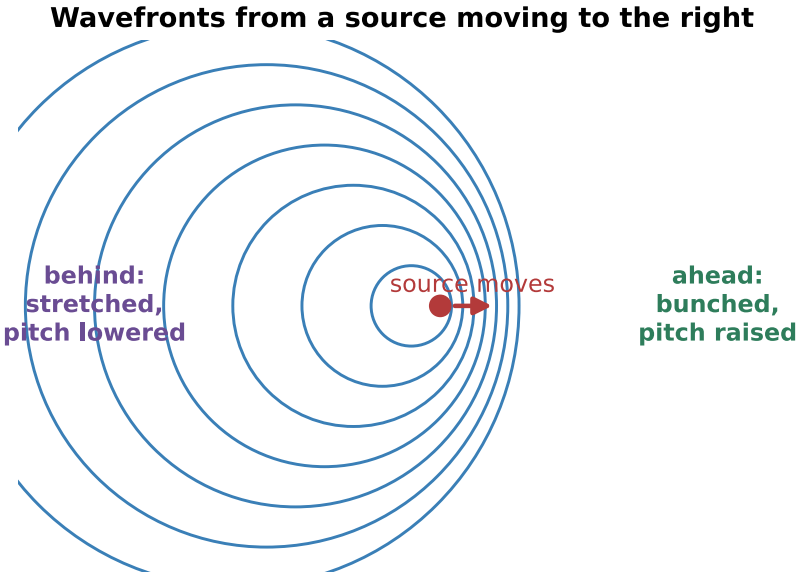


Figure 2.9: Wavefronts from a moving source. The wave speed is unchanged, every circle expands at the same rate, but the crests pile up ahead of the source and spread out behind it. A listener ahead meets more crests per second and hears a higher pitch. (The source here is drawn moving at 0.7 of the speed of sound so that the effect is visible; a passing car is nearer 0.09.)

Since frequency is crests per second, a listener ahead of the source hears a higher frequency and one behind hears a lower one:

$$f' = f\,\frac{v}{v \mp v_s}, \tag{2.12}$$

with the minus sign for a source approaching and the plus sign for one receding. If instead the *listener* moves and the source is still, the formula is $f' = f(v \pm v_l)/v$: a different expression, because the two cases are physically different. A moving source changes the wavelength in the air; a moving listener does not, and merely meets the existing crests at a different rate.

Worked example: an ambulance

A siren sounds at 700 Hz. The ambulance passes you at 25 m/s. What frequencies do you hear as it approaches and as it recedes? Take $v = 343$ m/s.

Approaching:

$$f' = (700\text{ Hz})\frac{343\text{ m/s}}{343\text{ m/s} - 25\text{ m/s}} = (700\text{ Hz})(1.0786) = 755\text{ Hz}. \tag{2.13}$$

Receding:

$$f' = (700\text{ Hz})\frac{343\text{ m/s}}{343\text{ m/s} + 25\text{ m/s}} = (700\text{ Hz})(0.9321) = 652\text{ Hz}. \tag{2.14}$$

The drop across the pass is $755 \to 652$ Hz, a ratio of 1.158, which is $1200\log_2(1.158) = 254$ cents, or about two and a half semitones. That is a musically large interval, which is why the effect is so obvious.

The Rotating Loudspeaker

The Doppler effect has one deliberate musical use, and it is a good one. A **Leslie speaker**, built for the Hammond organ, mounts a loudspeaker on a rotating baffle so that the source is continuously swinging toward and away from the listener. The result is a rich, swirling modulation of both pitch and loudness, quite unlike the flat vibrato of an electronic oscillator, and it became the characteristic sound of an entire era of organ playing.

What makes it sound organic rather than mechanical is that the modulation is not a simple oscillation. The pitch shift depends on the component of the source's velocity along the line to the listener, which varies as a sinusoid; the loudness depends on distance and on the direction the horn happens to be facing; and the room adds reflections that arrive with their own, different, modulations. The effect is a mixture of many slightly different versions of the same swirl.

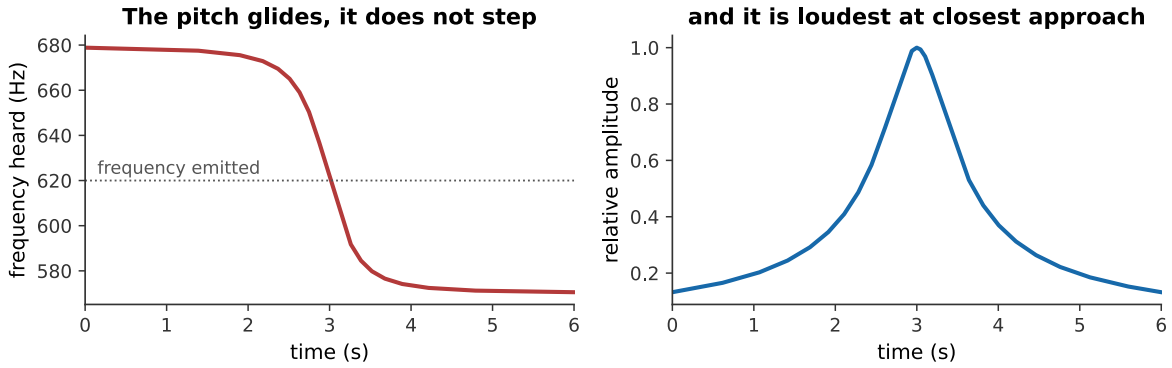


Figure 2.10: A source passing a listener at 30 m/s, about 110 km/h, missing by 12 m. The pitch does not step from high to low; it *glides*, and it glides fastest at the closest approach, because that is where the component of the velocity along the line of sight changes most rapidly. A two-tone caricature of a passing siren misses exactly this, which is why it never sounds convincing. *What you hear:*A tone approaching, passing, and receding: the pitch glides steadily downward, falling fastest at the moment of closest approach, and the loudness peaks at the same moment. Listen: Doppler Pass

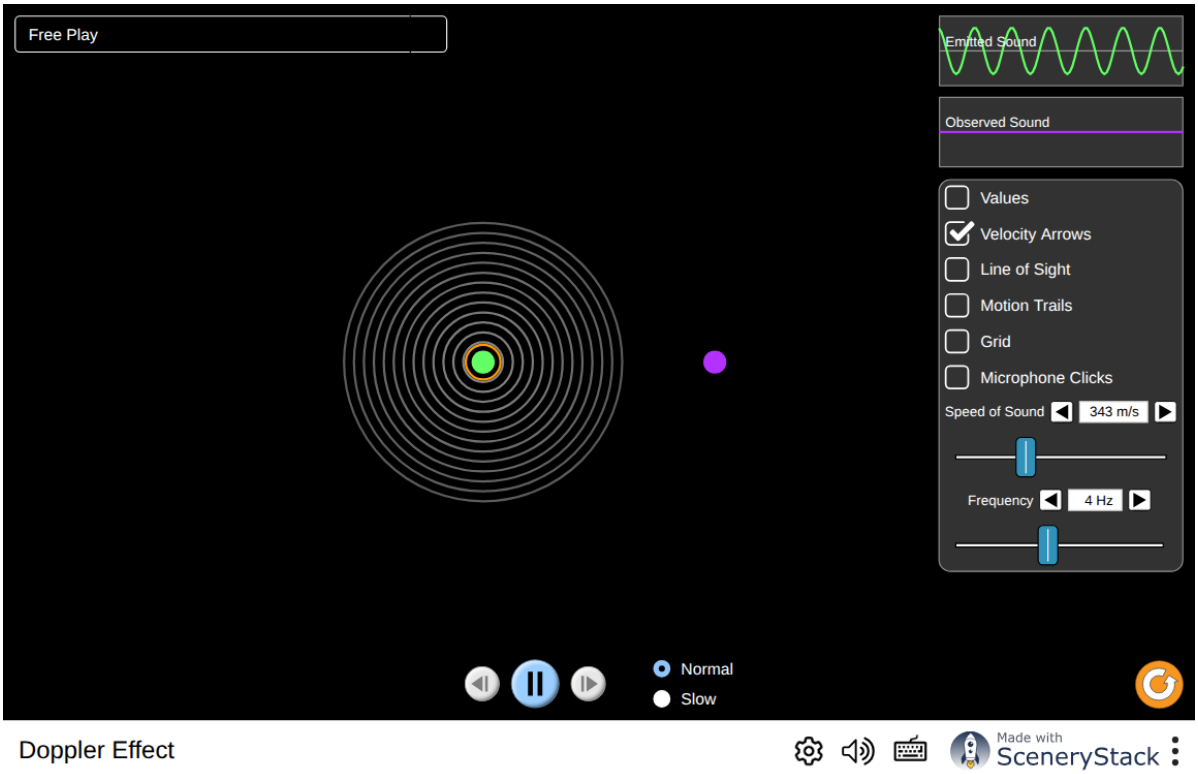


Figure 2.11: Move the source and the listener independently and watch the wavefronts. Two things are worth confirming deliberately: that the *wave speed* is unchanged in every case, and that moving the source and moving the listener at the same speed do not produce quite the same shift. Interactive simulation: Doppler Effect

Why a Passing Siren Falls and a Vibrato Does Not

A final observation that ties the section to music. A string player's **vibrato** rocks the finger back and forth, changing the string's sounding length and therefore its frequency, several times a second. The pitch really does go up and down. A Doppler shift would do something similar; so why does a violinist's vibrato not sound like a passing ambulance?

Two reasons. The vibrato's excursion is small, a few tens of cents rather than a few hundred, and it is *periodic*: it returns, several times a second, to the note it started from. The ear tracks the average and hears a single pitch with a shimmer on it. The siren's shift is large and monotonic, it goes down and stays down, so the ear has no average to settle on and hears the change itself. Chapter 8 takes up the question of what the auditory system does with a frequency that is moving.

2.6 Summary

- **A wave carries a disturbance, not matter.** Its speed is set by the medium, through $v = \sqrt{\text{stiffness}/\text{density}}$, and not by the source, the frequency, or the amplitude.
- **Transverse waves move the medium across the direction of travel; longitudinal waves move it along.** Strings carry the first, air the second, and the handover between them is the central design problem of a string instrument.
- $v = f\lambda$ follows from the observation that a wave advances one wavelength per period. Musical wavelengths in air run from about 12 m to about 8 cm, comparable to rooms and to hands, so sound diffracts around everyday obstacles.
- **Crossing a boundary conserves frequency**, not wavelength: the far side is driven by the near side at whatever rate it is shaken. Speed and wavelength both change.
- **In air,** $v = 331.3 + 0.606\,T$ **m/s**, depending on temperature and essentially nothing else, not on pressure, and only slightly on humidity. A fixed-length pipe tuned at 20 °C is about 46 cents flat at 5 °C, so winds and strings drift apart as a hall warms.
- **Intensity falls as** $1/r^2$ **in the open air**, 6 dB per doubling of distance, because the same power is spread over a growing sphere. Indoors this fails beyond the critical distance, where reflected sound dominates.
- **Instruments become more directional as pitch rises**, because directivity is governed by source size compared with wavelength, and only the wavelength changes.
- **Acoustic impedance governs how much sound crosses a boundary.** The mismatch between solids and air is huge, so a bare string radiates almost nothing; soundboards, bells, and the middle ear all exist to bridge that gap.
- **The Doppler effect** shifts the frequency of a moving source to $f' = fv/(v \mp v_s)$, by bunching or stretching the wavefronts. A moving listener gives a different formula, because the physics is different.

2.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity. The set mixes multiple response, true/false, drag-the-words, fill-in-the-blank, and mark-the-words formats, and one question reads its answer off the speed-of-sound chart.

1. **Multiple response.** A sound wave passes from air into water. Which of its properties change at the boundary? Select from: its speed; its wavelength; its frequency; its period.

2. **Visual true or false.** The activity reproduces the speed-of-sound bar chart from earlier in this chapter, which shows sound traveling nearly three times faster in helium than in air. The reason is that helium is far stiffer than air.

3. **Drag the words.** Complete: sound travels at about _____ m/s in air at 20 °C, gaining roughly _____ m/s per degree Celsius; it does not depend on the _____ of the air, nor on the _____ of the note. (Terms: 343, 0.6, pressure, frequency, 965, amplitude.)

4. **Fill in the blanks.** Doubling the distance from a small source in the open air lowers the level by _____ dB, because intensity falls as one over the distance _____.

5. **Mark the words.** Identify the two properties of the medium that set a wave's speed: "Wave speed is fixed by the medium's stiffness and its density; the frequency, the amplitude, and the motion of the source all leave it unchanged."

2.8 Conceptual Questions

1. A trumpeter thirty meters away plays a note. Explain why you hear it a tenth of a second later even though no air travels from the trumpet to your ear.

2. Explain why the speed of sound in air depends on temperature but not on atmospheric pressure, given that both affect how tightly packed the air is.

3. A sound wave passes from air into water. State which of frequency, wavelength, and speed change, and justify the one that does not.

4. A soundboard is cut so that the grain runs along the length of the instrument. Explain the acoustic reason, using the figures for spruce in §2.3.

5. Two listeners stand 2 m and 8 m from a singer, outdoors. By how many decibels do their levels differ? Explain why the same pair of positions in a small room would differ by much less.

6. Explain why a trumpet sounds much brighter to a listener directly in front of it than to one standing beside it, and why the effect is stronger for high notes.

7. A vibrating tuning fork is nearly silent in the air but loud when its base touches a table. Explain this using impedance, and say where the energy goes in each case.

8. A police car passes you at constant speed with its siren on. Sketch the pitch you hear against time, and explain why the steepest part of the curve occurs at the moment the car is closest, rather than before or after.

2.9 Problems

Exercise 2.1
(Straightforward) The speed of sound in air at 20 °C is 343 m/s. Find the wavelength of (a) a 100 Hz note, (b) a 1000 Hz note, (c) a 10,000 Hz note.

Solution to Exercise 2.1
Using $\lambda = v/f$ throughout:

$$\lambda_{100} = \frac{343\ \text{m/s}}{100\ \text{Hz}} = 3.43\ \text{m}, \qquad \lambda_{1000} = \frac{343\ \text{m/s}}{1000\ \text{Hz}} = 0.343\ \text{m}, \qquad \lambda_{10000} = \frac{343\ \text{m/s}}{10000\ \text{Hz}} = 0.0343\ \text{m}. \tag{2.15}$$

Therefore, the wavelengths are 3.43 m, 34.3 cm, and 3.43 cm. Each factor of ten in frequency is a factor of ten in wavelength, in the opposite direction.

Exercise 2.2
(Straightforward) Find the speed of sound in air at (a) 0 °C, (b) 25 °C, (c) -15 °C.

Solution to Exercise 2.2
Using $v = 331.3 + 0.606\,T$:

$$v(0) = 331.3\ \text{m/s}, \qquad v(25) = 331.3{+}15.2 = 346.5\ \text{m/s}, \qquad v(-15) = 331.3{-}9.1 = 322.2\ \text{m/s}. \tag{2.16}$$

Therefore, the speeds are 331, 347, and 322 m/s: a spread of about 7% over a range of temperatures a marching band might actually meet.

Exercise 2.3
(Straightforward) You see a lightning flash and hear the thunder 4.2 s later. The air temperature is 18 °C. How far away was the strike? (Light's travel time is negligible.)

Solution to Exercise 2.3
First the speed:

$$v = 331.3 + 0.606(18) = 342.2\ \text{m/s}. \tag{2.17}$$

Then the distance:

$$d = vt = (342.2\ \text{m/s})(4.2\ \text{s}) = 1.44 \times 10^{3}\ \text{m}. \tag{2.18}$$

Therefore, the strike was about 1.4 km away. The familiar rule of thumb, three seconds per kilometer, follows from $1000/342 = 2.9$.

Exercise 2.4
(Moderate) An organ pipe is tuned to sound A440 in a church at 18 °C. The heating fails and the church cools to 6 °C. (a) What frequency does the pipe now sound? (b) Express the change in cents. (c) Would a listener notice?

Solution to Exercise 2.4
(a) The pipe's length is fixed, so its frequency is proportional to the speed of sound:

$$v(18) = 342.2\ \text{m/s}, \qquad v(6) = 331.3 + 3.64 = 334.9\ \text{m/s}, \tag{2.19}$$

$$f' = (440\ \text{Hz})\frac{334.9\ \text{m/s}}{342.2\ \text{m/s}} = 430.6\ \text{Hz}. \tag{2.20}$$

(b) In cents:

$$n = 1200 \log_2\!\left(\frac{430.6}{440}\right) = 1200 \times (-0.0312) = -37 \text{ cents}. \tag{2.21}$$

(c) Yes, unmistakably. The just-noticeable difference for a trained listener comparing two sustained tones is a few cents; 37 cents is more than a third of a semitone, and any other instrument playing alongside would clash badly.

Exercise 2.5
(Straightforward) A guitar string of length 65 cm has mass per unit length $\mu = 6.3 \times 10^{-3}$ kg/m and is under tension $T = 110$ N. (a) What is the wave speed along the string? (b) The lowest mode has a wavelength twice the string length. What frequency does the string sound?

Solution to Exercise 2.5
(a) Using $v = \sqrt{T/\mu}$:

$$v = \sqrt{\frac{110 \text{ N}}{6.3 \times 10^{-3} \text{ kg/m}}} = \sqrt{1.746 \times 10^{4}} = 132 \text{ m/s}. \tag{2.22}$$

(b) With $\lambda = 2L = 1.30$ m:

$$f = \frac{v}{\lambda} = \frac{132 \text{ m/s}}{1.30 \text{ m}} = 102 \text{ Hz}. \tag{2.23}$$

Therefore, waves travel at 132 m/s along the string and it sounds about 102 Hz, close to the G_2 of a guitar's third string. Note how much *slower* this is than sound in air: the wave on the string and the sound it radiates have quite different speeds and therefore quite different wavelengths, though they share a frequency.

Exercise 2.6
(Straightforward) A tone of 440 Hz travels from air ($v = 343$ m/s) into water ($v = 1482$ m/s). Find its wavelength in each medium, and state its frequency in the water.

Solution to Exercise 2.6
In air:

$$\lambda_{\text{air}} = \frac{343 \text{ m/s}}{440 \text{ Hz}} = 0.780 \text{ m}. \tag{2.24}$$

Frequency is conserved across the boundary, so in water it is still 440 Hz, and

$$\lambda_{\text{water}} = \frac{1482 \text{ m/s}}{440 \text{ Hz}} = 3.37 \text{ m}. \tag{2.25}$$

Therefore, the wavelength grows by the ratio of the speeds, $1482/343 = 4.32$, while the pitch is unchanged.

Exercise 2.7
(Moderate) A busker plays outdoors. A listener 3.0 m away measures a sound intensity of 4.0×10^{-5} W/m^2. (a) What is the busker's acoustic power output, assuming uniform radiation? (b) What intensity would a listener 12 m away measure?

Solution to Exercise 2.7

(a) From $I = P/4\pi r^2$:

$$P = 4\pi r^2 I = 4\pi(3.0\ \text{m})^2(4.0\times10^{-5}\ \text{W/m}^2) = 4.5\times10^{-3}\ \text{W}. \tag{2.26}$$

(b) Four times the distance is sixteen times the area:

$$I = \frac{4.0\times10^{-5}\ \text{W/m}^2}{16} = 2.5\times10^{-6}\ \text{W/m}^2. \tag{2.27}$$

Therefore, the busker radiates about 4.5 mW of sound, and the distant listener receives 2.5 μW/m^2. Note how small the power is: a few milliwatts is a perfectly audible musical performance, which says more about the sensitivity of the ear than about the efficiency of the busker.

Exercise 2.8

(Moderate) A train sounds a 520 Hz horn and approaches a station platform at 31 m/s. Take $v = 343$ m/s. (a) What frequency does a waiting passenger hear as it approaches? (b) As it recedes? (c) Express the total change in cents.

Solution to Exercise 2.8

(a) Approaching:

$$f' = (520\ \text{Hz})\frac{343\ \text{m/s}}{343\ \text{m/s} - 31\ \text{m/s}} = (520\ \text{Hz})(1.0994) = 572\ \text{Hz}. \tag{2.28}$$

(b) Receding:

$$f' = (520\ \text{Hz})\frac{343\ \text{m/s}}{343\ \text{m/s} + 31\ \text{m/s}} = (520\ \text{Hz})(0.9171) = 477\ \text{Hz}. \tag{2.29}$$

(c) In cents:

$$n = 1200\log_2\!\left(\frac{572}{477}\right) = 1200(0.2622) = 315\ \text{cents}. \tag{2.30}$$

Therefore, the horn falls from 572 Hz to 477 Hz, a drop of 315 cents: a little over three semitones, roughly a minor third.

Exercise 2.9

(Moderate) Sound travels at 5960 m/s in steel and 343 m/s in air. A steel pipe 180 m long is struck at one end. (a) Find the two arrival times at the far end. (b) A listener claims to hear only one sound. Suggest the most likely reason.

Solution to Exercise 2.9

(a) Through the steel and through the air:

$$t_{\text{steel}} = \frac{180\ \text{m}}{5960\ \text{m/s}} = 0.0302\ \text{s}, \qquad t_{\text{air}} = \frac{180\ \text{m}}{343\ \text{m/s}} = 0.525\ \text{s}. \tag{2.31}$$

The gap is 0.49 s.

(b) Half a second is ample separation, so a listener hearing one sound is most likely not in contact with the pipe: the steel-borne pulse has to be coupled into the listener, by touch or by the pipe radiating at the far end, and if the pipe is well isolated at that end the steel-borne arrival may be too faint to notice.

Exercise 2.10
(Moderate) A loudspeaker 20 cm across is used to reproduce a range of frequencies. Estimate the frequency above which the speaker's diameter exceeds one wavelength, and comment on what happens to its radiation pattern above that frequency.

Solution to Exercise 2.10
Setting $\lambda = 0.20$ m and solving for frequency:

$$f = \frac{v}{\lambda} = \frac{343\ \text{m/s}}{0.20\ \text{m}} = 1.7 \times 10^{3}\ \text{Hz}. \tag{2.32}$$

Therefore, above roughly 1.7 kHz the driver is larger than a wavelength. Below that it radiates nearly uniformly into the room; above it, different parts of the cone interfere and the output narrows into a beam along the axis, so the listener hears progressively less treble as they move off-axis. This is precisely why loudspeakers use a small tweeter for high frequencies rather than one driver for everything (Chapter 15).

Exercise 2.11
(Moderate) A flute is tuned to A440 in a 21 °C rehearsal room, then carried outdoors to play at 9 °C. (a) What frequency does it sound outdoors, before the player compensates? (b) A flutist can compensate by rolling the instrument and adjusting the embouchure over a range of about ±25 cents. Is that enough?

Solution to Exercise 2.11
(a) The speeds are

$$v(21) = 331.3 + 12.7 = 344.0\ \text{m/s}, \qquad v(9) = 331.3 + 5.5 = 336.8\ \text{m/s}, \tag{2.33}$$

so

$$f' = (440\ \text{Hz})\frac{336.8\ \text{m/s}}{344.0\ \text{m/s}} = 430.8\ \text{Hz}. \tag{2.34}$$

(b) In cents:

$$n = 1200 \log_2\!\left(\frac{430.8}{440}\right) = -37\ \text{cents}. \tag{2.35}$$

Therefore, the flute plays 37 cents flat, which exceeds the ±25 cent range the player can cover. The physical remedy is to shorten the instrument by pushing the head joint in, which is exactly what flutists do when the temperature changes.

Exercise 2.12
(Challenging) Two microphones on a line are 4.00 m apart. A hand clap made beyond one of them, on that line, arrives at the second microphone 11.9 ms after the first. (a) What speed of sound does this imply? (b) What air temperature does that correspond to? (c) If the clap were actually made 1.0 m off the line, would the inferred speed be too high or too low? Explain.

Solution to Exercise 2.12

(a) From speed as distance over time:

$$v = \frac{4.00\ \text{m}}{11.9 \times 10^{-3}\ \text{s}} = 336\ \text{m/s}. \tag{2.36}$$

(b) Solving $336 = 331.3 + 0.606\,T$:

$$T = \frac{336 - 331.3}{0.606} = 7.8^\circ\,\text{C}. \tag{2.37}$$

(c) Too high. Off the line, the *difference* in path lengths to the two microphones is less than their 4.00 m separation, so the measured delay is shorter than it should be for that separation. Dividing the full 4.00 m by that too-short delay overestimates the speed. This is the systematic error the laboratory exercise in Laboratory Exercises warns about.

Exercise 2.13

(Challenging) Helium's speed of sound is 965 m/s, against 343 m/s for air. A singer inhales helium and sings a sustained vowel. (a) By what factor do the resonances of their vocal tract shift? (b) Does the pitch of the note change? Explain carefully, and say what the listener actually hears differently.

Solution to Exercise 2.13

(a) The vocal tract's resonances are set by its length and by the speed of sound in the gas filling it. The length is unchanged, so the resonant frequencies scale directly with the speed:

$$\frac{965}{343} = 2.81. \tag{2.38}$$

(b) **No, the pitch is essentially unchanged.** The pitch is set by the rate at which the vocal folds open and close, which is governed by their mass and tension and by the air pressure driving them, not by the speed of sound in the gas. What shifts is the *filter*, not the *source*.

Therefore, the listener hears the same note with its resonances moved up by nearly a factor of three, which strips the low-frequency emphasis out of the vowel and leaves the characteristic thin, quacking timbre. This is the clearest everyday demonstration of the source–filter model of Chapter 13, and the popular description of helium as "raising your voice" gets the physics exactly backwards.

Exercise 2.14

(Moderate) A Leslie speaker's horn rotates on a circle of radius 0.17 m at 6.6 revolutions per second. (a) What is the speed of the horn's mouth? (b) What is the maximum Doppler shift, in cents, for a 500 Hz tone heard by a distant listener? (c) How many times per second does the listener hear the pitch rise and fall?

Solution to Exercise 2.14

(a) The speed is the circumference divided by the period:

$$v_s = 2\pi r f_{\text{rot}} = 2\pi(0.17\ \text{m})(6.6\ \text{s}^{-1}) = 7.05\ \text{m/s}. \tag{2.39}$$

(b) The extreme shifts occur when the horn moves directly toward and directly away from the listener:

$$f'_{\text{near}} = (500\,\text{Hz})\frac{343\ \text{m/s}}{343\ \text{m/s} - 7.05\ \text{m/s}} = 510.5\,\text{Hz}, \qquad f'_{\text{far}} = (500\,\text{Hz})\frac{343\ \text{m/s}}{343\ \text{m/s} + 7.05\ \text{m/s}} = 489.9\,\text{Hz}. \tag{2.40}$$

In cents, relative to 500 Hz:

$$1200\log_2\left(\frac{510.5}{500}\right) = +36\ \text{cents}, \qquad 1200\log_2\left(\frac{489.9}{500}\right) = -35\ \text{cents}. \tag{2.41}$$

(c) Once per revolution, so 6.6 times per second.

Therefore, the pitch swings about ± 36 cents at 6.6 Hz: a wide, fast vibrato, which together with the accompanying loudness modulation is the characteristic Leslie swirl.

Chapter 3

Superposition, Interference, and Standing Waves

Learning Objectives

By the end of this chapter, you should be able to:

- State the superposition principle, apply it graphically to find the resultant of two overlapping waves, and identify the condition under which it fails.
- Convert a path-length difference into a phase difference, and predict whether two waves from separated sources interfere constructively or destructively at a given point.
- Explain why two loudspeakers playing the same tone produce dead spots in a room, and estimate the spacing between them.
- Describe what happens to a wave pulse reflected from a fixed end and from a free end, and explain the inversion at a fixed end.
- Explain how a standing wave arises from two identical waves traveling in opposite directions, and locate its nodes and antinodes.
- Derive the frequencies of the standing-wave modes of a string fixed at both ends, $f_n = nv/2L$, and use them to explain why a shorter or tighter string sounds higher.
- Derive the mode frequencies of an open and of a stopped air column, and explain why a stopped pipe sounds an octave lower than an open pipe of the same length and produces only odd harmonics.
- Explain why a standing wave is the reason an instrument has a definite pitch at all.

Introduction

So far the book has followed one wave at a time. Music never does. A room is full of waves, the direct sound from the instrument, the reflection off the back wall, the note the second violin is playing, the reflection of *that* off the ceiling, all arriving at your ear at once and all adding up.

What they add up to is the subject of this chapter, and it is governed by one very simple rule with one very surprising consequence.

The rule is **superposition**: where two waves overlap, the displacements simply add. There is no interaction, no collision, no scattering of one wave off the other. Two waves pass through each other and emerge unchanged, as though nothing had happened.

The consequence is that waves can cancel. Two sounds can arrive at a point and produce silence there. And when a wave meets a wall, reflects, and runs back into itself, the adding-and-canceling settles into a stationary pattern, a **standing wave**, that can only exist at certain

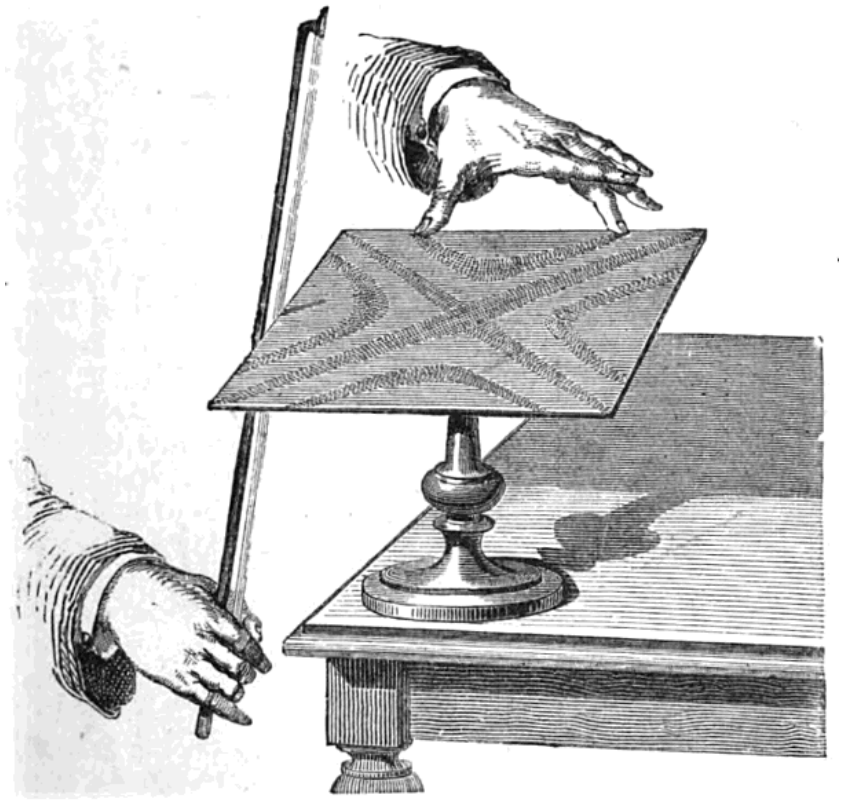


Figure 3.1: An early experiment in making modes visible: a bowed plate gathers sand along its nodal lines. Public-domain illustration from 1879.

frequencies.

That last sentence is the reason musical instruments have pitches. Everything in Part V is an elaboration of it.

3.1 The Superposition Principle

Adding Waves Point by Point

The rule is as simple as it sounds. If wave 1 alone would produce a displacement y_1 at some point and some moment, and wave 2 alone would produce y_2 there and then, the two together produce

$$y = y_1 + y_2. \tag{3.1}$$

That is all. The waves do not know about each other.

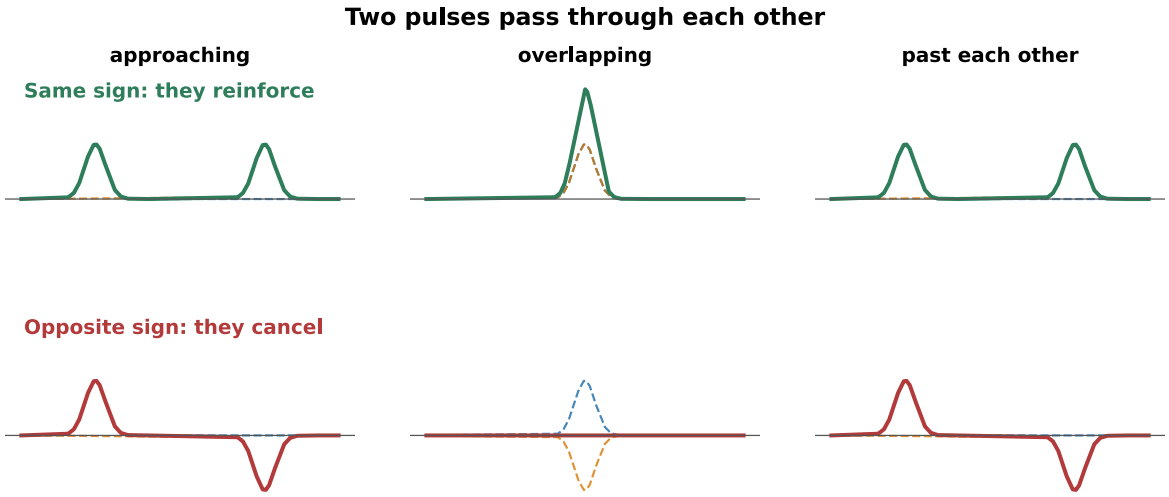


Figure 3.2: Two pulses passing through each other. In the top row they have the same sign and momentarily reinforce; in the bottom row they have opposite signs and momentarily cancel. In both cases they emerge on the other side exactly as they went in. The cancellation in the middle of the bottom row is complete, at that instant the medium is flat, but it is not the end of the pulses.

The bottom row of that figure deserves a moment's thought, because it looks like a violation of energy conservation. At the instant the two pulses exactly cancel, the string is perfectly straight. Where has the energy gone?

Into motion. The string is straight but it is not still: every point of it is moving, fast, and the energy is entirely kinetic at that instant. A moment later the pulses re-emerge. Superposition adds *displacements*, and displacement is not the only place a wave keeps its energy.

Constructive and Destructive Interference

When two waves of the same frequency overlap steadily, what matters is their relative **phase**.

- In step, crest on crest, they reinforce. This is **constructive interference**, and the resulting amplitude is the sum of the two.

- Exactly out of step, crest on trough, they cancel. This is **destructive interference**, and if the amplitudes are equal the result is nothing at all.
- In between, something in between.

Two loudspeakers playing the same steady tone are the standard demonstration. Walk across the room in front of them and the tone rises and falls: loud where the two arrivals are in step, and startlingly near-silent where they are opposed.

> **Warning**
> Destructive interference does not destroy energy; it moves it. The energy missing from a quiet spot is in the loud spots, which are *four* times as intense as either source alone would produce, not twice. Adding amplitudes and then squaring is not the same as adding intensities, and the difference is exactly what interference is.

Where Superposition Breaks Down

Superposition holds because the wave equation is **linear**: the restoring force is proportional to the displacement, as in Chapter 1. Everything in this chapter depends on that.

It is an approximation, and it fails at large amplitudes. In an extremely loud sound wave the compressions travel slightly faster than the rarefactions, because compressed air is hotter and therefore stiffer; a wave that starts sinusoidal steepens as it travels, and at the limit it becomes a shock. This is not a musical regime, it takes something like a jet engine or an explosion, but it lurks in two places the book will return to: inside a brass instrument played very loudly, where the wave really does steepen on its way down the bore and adds a characteristic brightness to a fortissimo (Chapter 11), and in a cymbal driven hard, where nonlinearity spills energy into a dense forest of new frequencies (Chapter 12).

3.2 Interference in Space

Path Difference and Phase Difference

The phase relationship between two arrivals at a listener is set by how far each has traveled. If the two sources emit in step, then all that matters is the **path difference**

$$\Delta r = r_2 - r_1. \tag{3.2}$$

A path difference of exactly one wavelength puts the two arrivals back in step: the second has been delayed by exactly one full cycle. A path difference of half a wavelength puts them exactly opposed. In general, the phase difference is

$$\phi = \frac{2\pi}{\lambda}\,\Delta r, \tag{3.3}$$

and the conditions are:

$$\text{loud (constructive):}\quad \Delta r = m\lambda, \qquad \text{quiet (destructive):}\quad \Delta r = \left(m + \tfrac{1}{2}\right)\lambda, \tag{3.4}$$

with $m = 0, 1, 2, \ldots$

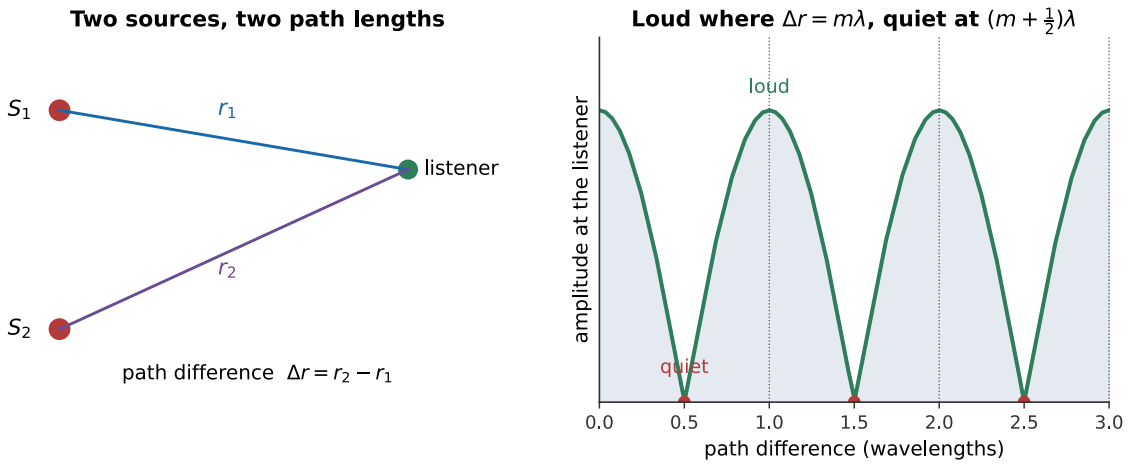


Figure 3.3: Two sources and a listener sweeping past them. Everything depends on the difference between the two path lengths, measured in wavelengths, tracked by the moving dot on the right. The listener hears maximum loudness whenever that difference is a whole number of wavelengths and silence whenever it is a half-integer.

Worked example: finding a dead spot

Two loudspeakers, 2.4 m apart, both play a steady 860 Hz tone in step. A listener stands 4.0 m from one speaker. How far from the other must they be to hear nothing? Take $v = 343$ m/s.

The wavelength is

$$\lambda = \frac{v}{f} = \frac{343 \text{ m/s}}{860 \text{ Hz}} = 0.399 \text{ m}. \tag{3.5}$$

Silence requires a path difference of half a wavelength, or one and a half, or two and a half. The smallest is

$$\Delta r = \tfrac{1}{2}\lambda = 0.199 \text{ m}, \tag{3.6}$$

so the listener must be 4.0 ± 0.20 m from the other speaker: that is, 4.20 m or 3.80 m.

Notice how *small* that displacement is. Moving your head twenty centimeters can take you from full loudness to near-silence at this frequency, which is why the effect is so easy to demonstrate and so disconcerting the first time.

Two Sources in a Room

The worked example understates the problem in a real room, because a real room has far more than two sources. Every reflecting surface acts as an image of the loudspeaker, so a single speaker in a rectangular room behaves like a large array of them, all radiating the same signal from different distances.

The result is that the sound field in a small room at low frequencies is a lumpy landscape of loud and quiet regions that does not change with time. Move your head and the bass changes. This is the room-mode problem, and Chapter 14 shows that it is worst exactly where it is most annoying: at low frequencies, where wavelengths are comparable to the room dimensions.

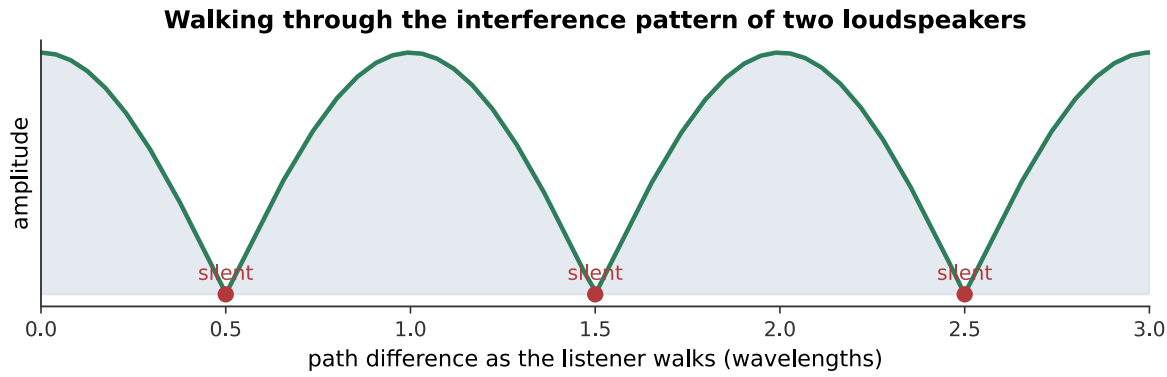


Figure 3.4: What a listener hears walking through the interference pattern of two loudspeakers playing the same 500 Hz tone. The loudspeakers are not changing; the *listener's position* is, and with it the path difference. Each near-silence is a half-integer number of wavelengths. *What you hear:*A steady tone that fades almost completely to silence three times, then returns to full loudness, as though someone were walking slowly past two loudspeakers. Listen: Walking Through Interference

Interference You Can Walk Through

Two things about that demonstration are worth drawing out, because both bear on music.

It is frequency-dependent. The pattern's spacing scales with the wavelength, so a given pair of sources produces a coarse pattern for bass and an extremely fine one for treble. Play *music* through two speakers rather than a single tone and no listener hears silence, because the nulls for different frequencies are in different places. What they hear instead is a position-dependent coloring of the sound, some frequencies reinforced, others suppressed, which is precisely the "comb filtering" that recording engineers work to avoid.

It requires a stable phase relationship. Two speakers fed from the same source stay in step indefinitely, so the pattern stands still. Two *different* instruments playing the same written note do not: their frequencies differ slightly and continuously, so the pattern drifts rather than standing still, and the listener hears beats (Chapter 8) rather than a fixed geography of loud and quiet.

3.3 Reflection and the Standing Wave

Reflection at a Fixed and a Free End

A wave traveling along a string reaches the end. What happens next depends entirely on what is there.

If the end is **clamped**, it cannot move. But the arriving pulse is trying to move it, so the clamp must push back with an equal and opposite force, and the reflected pulse comes back **inverted**.

If the end is **free**, attached to a ring that can slide on a frictionless rod, say, nothing resists the motion. The end overshoots, and the pulse comes back **the same way up**.

The same distinction applies to air columns, with one twist that catches everyone out. The **closed** end of a pipe is a place where the air cannot move: a displacement node, but it is a place where the pressure swings most: a pressure antinode. The **open** end is the reverse. Section 3.5 returns to this; you need to decide now which of the two quantities you are drawing, because the two pictures look like each other's opposites.

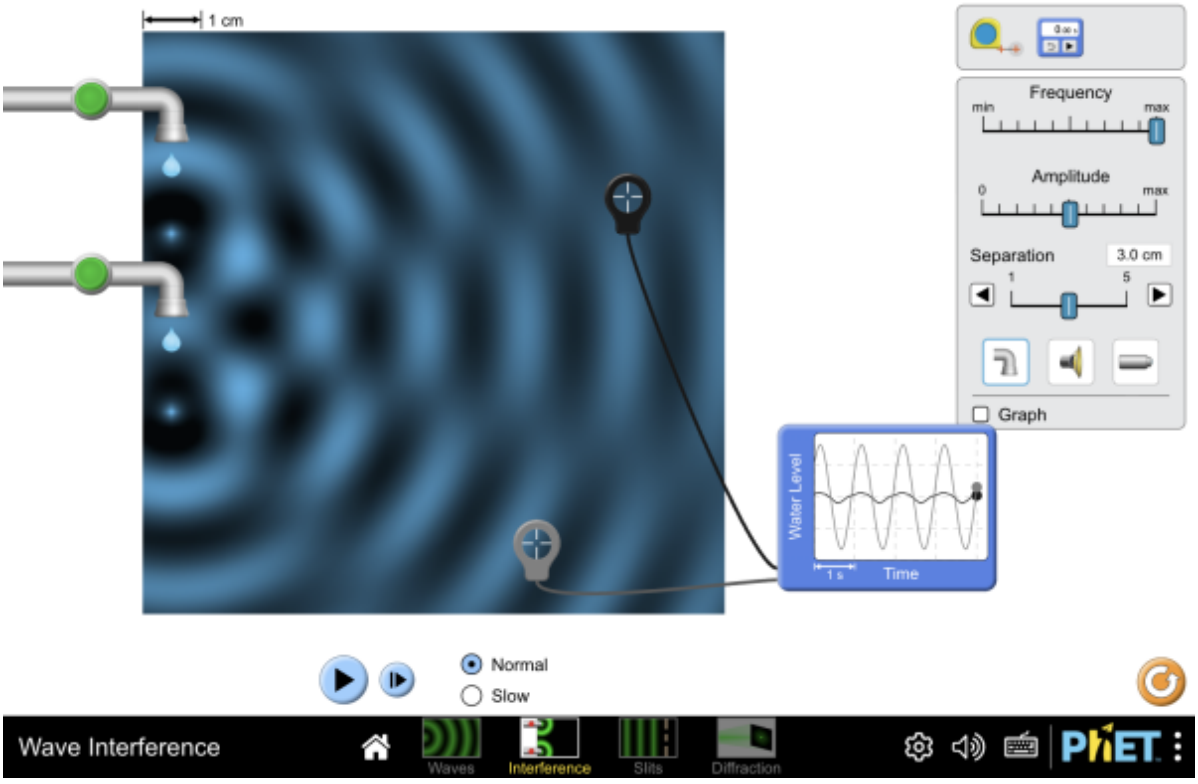


Figure 3.5: Two sources on the sound screen, with the interference pattern drawn out. Raise the frequency and note that the pattern's spacing narrows, the geometry is fixed, and only the wavelength has changed. The water screen shows the identical geometry with ripples instead of sound, if a visible medium helps first. Interactive simulation: Wave Interference

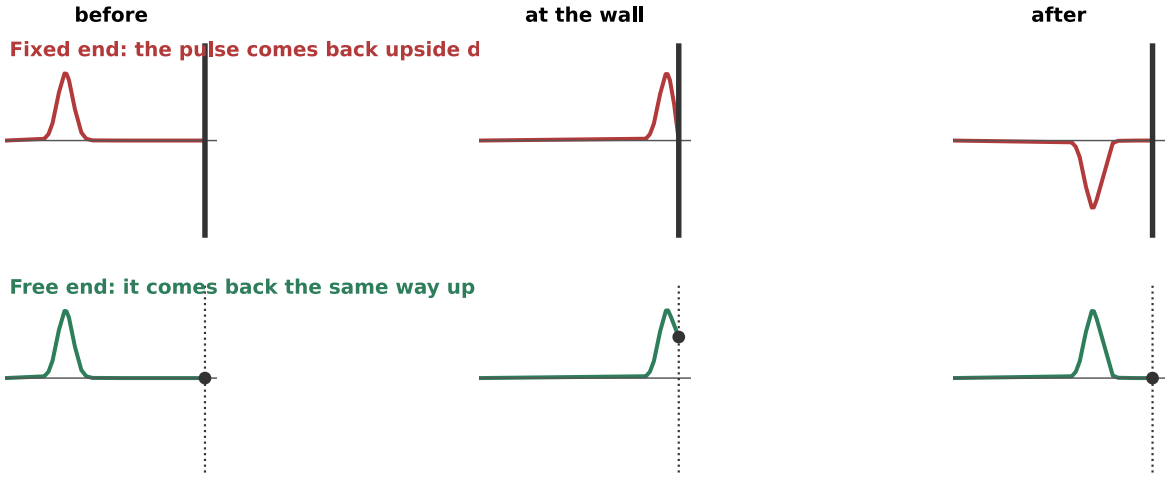


Figure 3.6: Reflection at the two extreme boundaries. A fixed end forces the displacement to zero there, which can only be arranged if the reflected pulse is inverted; a free end imposes no such condition and the pulse returns upright. Real boundaries lie between these extremes, and the fraction of energy reflected is governed by the impedance mismatch of §2.4.

Two Waves, Opposite Directions

Now put the two together. Send a continuous wave down a string toward a fixed end. It reflects and travels back, and from then on the string carries two identical waves traveling in opposite directions at once. Superposition says to add them.

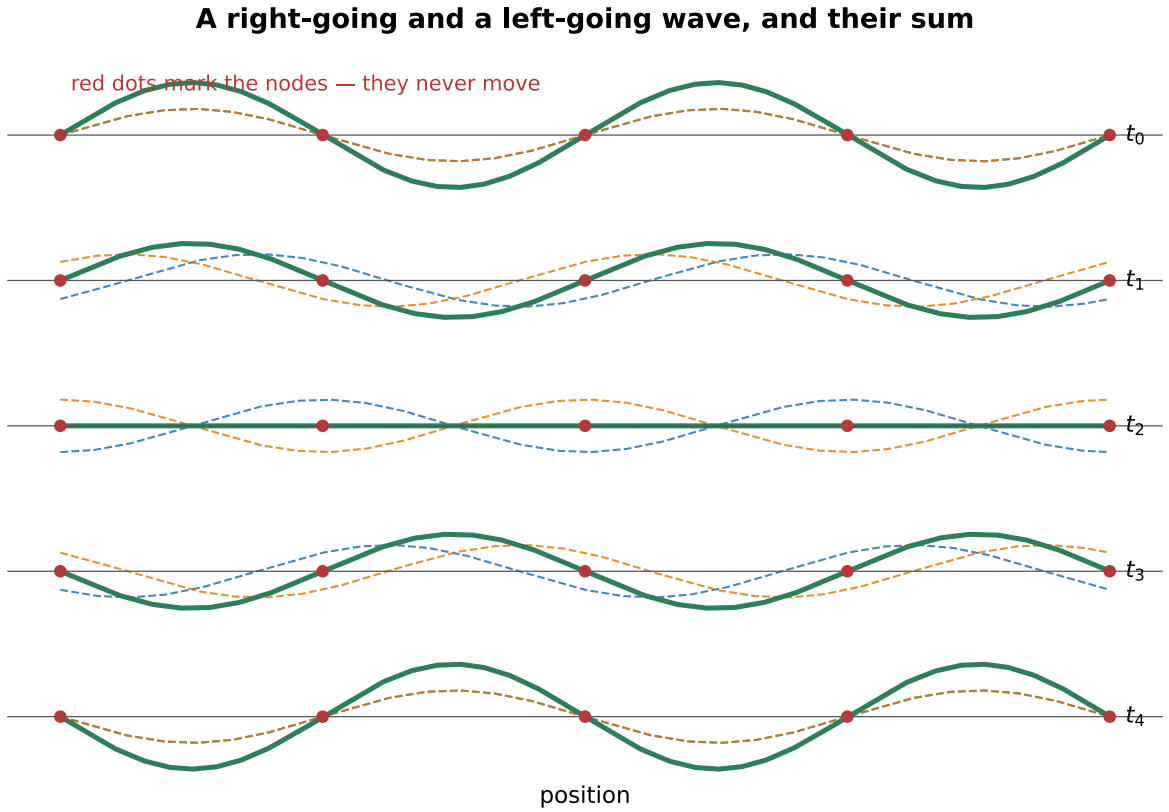


Figure 3.7: A right-going wave and a left-going wave of the same frequency and amplitude, added continuously. The sum does not travel. It stands still and breathes: the whole pattern grows, shrinks, inverts, and grows again, but the points where it crosses zero never move.

That is a **standing wave**, and the striking thing about it is that nothing propagates. Energy sloshes back and forth locally; the pattern itself goes nowhere.

The trigonometry, in two lines

Two identical waves traveling in opposite directions are $y_1 = A\sin(kx - \omega t)$ and $y_2 = A\sin(kx + \omega t)$. Adding them, and using $\sin P + \sin Q = 2\sin\frac{P+Q}{2}\cos\frac{P-Q}{2}$:

$$y = y_1 + y_2 = 2A\sin(kx)\cos(\omega t). \tag{3.7}$$

The result has factored: one factor depends only on position, the other only on time. That separation *is* the standing wave. The shape $\sin(kx)$ is fixed in space; the factor $\cos(\omega t)$ scales the whole of it up and down together.

Compare this with the traveling wave it came from, $\sin(kx - \omega t)$, in which x and t appear only in the combination $kx - \omega t$, which is what makes the pattern move. The difference between a wave that travels and a wave that stands is exactly whether space and time can be separated.

The nodes are where $\sin(kx) = 0$, that is $kx = m\pi$, or $x = m\lambda/2$. They are spaced half a wavelength apart, and they are *permanently* at zero: no value of t makes them move.

Nodes, Antinodes, and What Does Not Move

Two names for the parts of a standing wave:

- A **node** is a point that never moves. Nodes are half a wavelength apart.
- An **antinode** is a point that moves most. Antinodes lie halfway between nodes, also half a wavelength apart.

You can find the nodes of a guitar string by touching it very lightly at various points while it sounds. Touch an antinode and the note dies; touch a node and it carries on, because you are holding a point that was not moving anyway. This is not a party trick, it is exactly how a string player produces harmonics, and §3.4 uses it.

3.4 Standing Waves on a String

Fitting Half-Wavelengths Between Two Ends

A string fixed at both ends must have a node at each end. That is a boundary condition, and it is very restrictive: only those standing waves which happen to have nodes exactly L apart can exist on the string at all.

Since nodes are half a wavelength apart, the condition is that a whole number of half-wavelengths must fit into the length:

$$L = n\,\frac{\lambda_n}{2}, \qquad n = 1, 2, 3, \ldots \tag{3.8}$$

so

$$\lambda_n = \frac{2L}{n}, \qquad f_n = \frac{v}{\lambda_n} = n\,\frac{v}{2L}. \tag{3.9}$$

Combining this with the string's wave speed from Chapter 2, $v = \sqrt{T/\mu}$:

$$\boxed{f_n = \frac{n}{2L}\sqrt{\frac{T}{\mu}}} \tag{3.10}$$

This is the single most useful formula in the book's treatment of instruments, and every term in it is a lever an instrument maker or player can pull:

Change	Effect on pitch	Who uses it
Shorten L	Higher	Every fretting or stopping hand
Raise tension T	Higher	Tuning pegs
Raise mass per length μ	Lower	Wound bass strings

Note the **square roots**. Doubling the tension does not double the frequency; it raises it by a factor of $\sqrt{2}$, about seven semitones. To raise a string by an octave you must quadruple its tension; so instruments are tuned by shortening rather than by tightening, and the strings of a piano are of enormously varied length and mass rather than of varied tension alone.

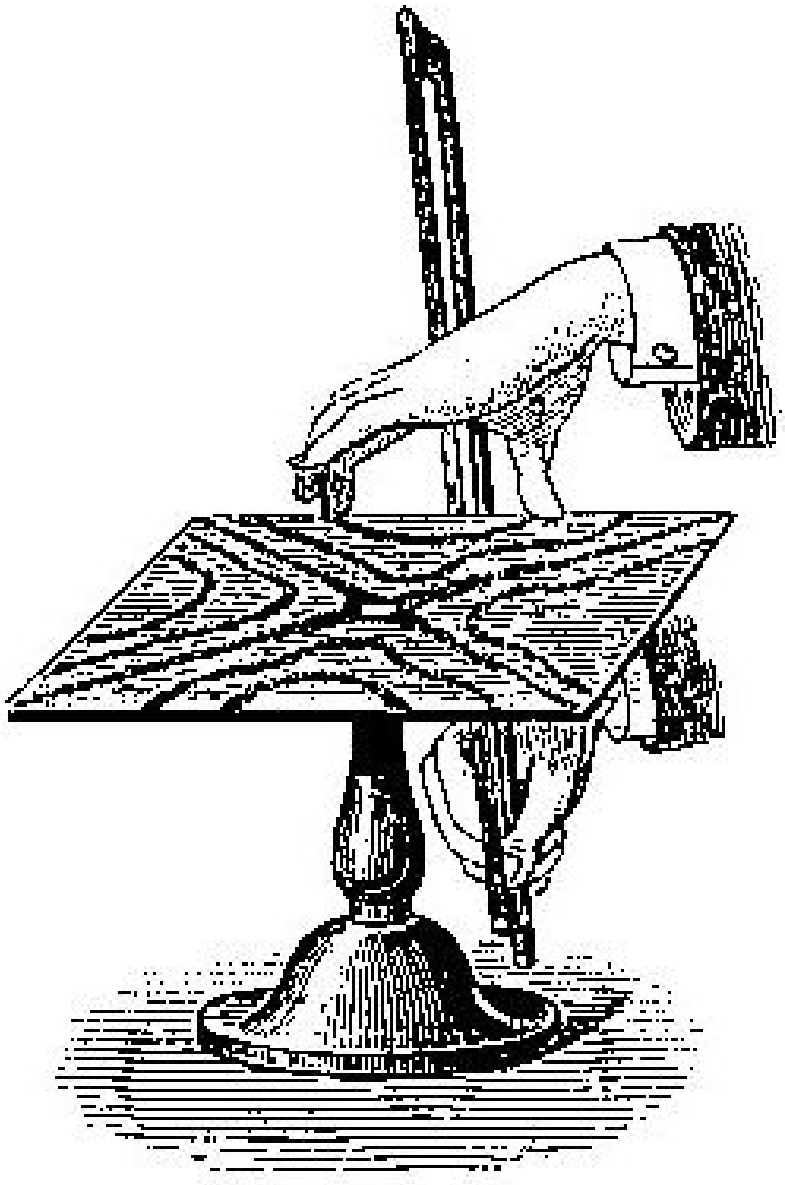


Figure 3.8: Nodal lines are not merely a drawing convention: powder makes them visible on a vibrating surface. Émile Desbeaux, public domain.

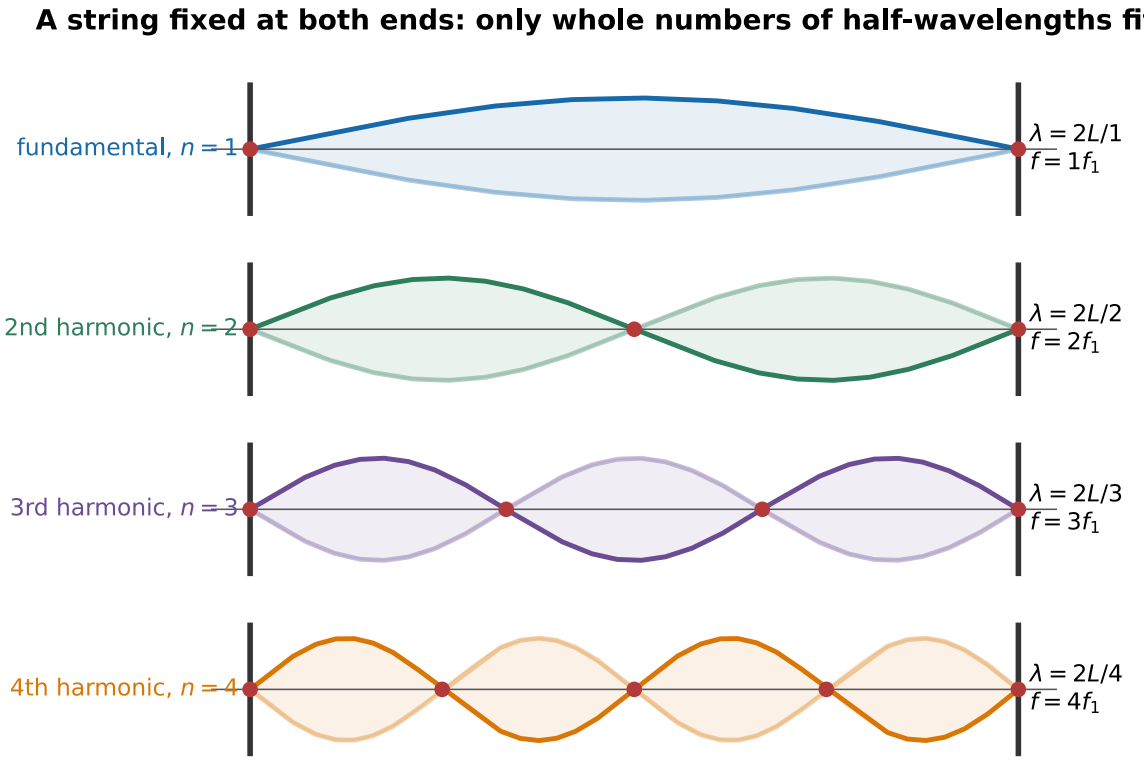


Figure 3.9: The first four modes of a string fixed at both ends, each drawn as an envelope the string sweeps between, with the instantaneous shape oscillating inside it. The nth mode has n half-wavelengths in the length, $n - 1$ nodes between the ends, and a frequency n times the fundamental — watch the higher modes oscillate faster.

The Harmonic Series of a String

The set of frequencies $f_n = nf_1$ is the **harmonic series**, and the fact that it comes out as exact whole-number multiples is not an accident of the algebra. It is a consequence of the boundary conditions being the same at both ends and of the wave speed being the same at every frequency.

It is also, arguably, the reason Western music sounds the way it does. Chapters 5, 8, and 9 are all downstream of it: the harmonic series determines what a plucked string's spectrum contains, which determines which intervals sound smooth, which determines what a scale is.

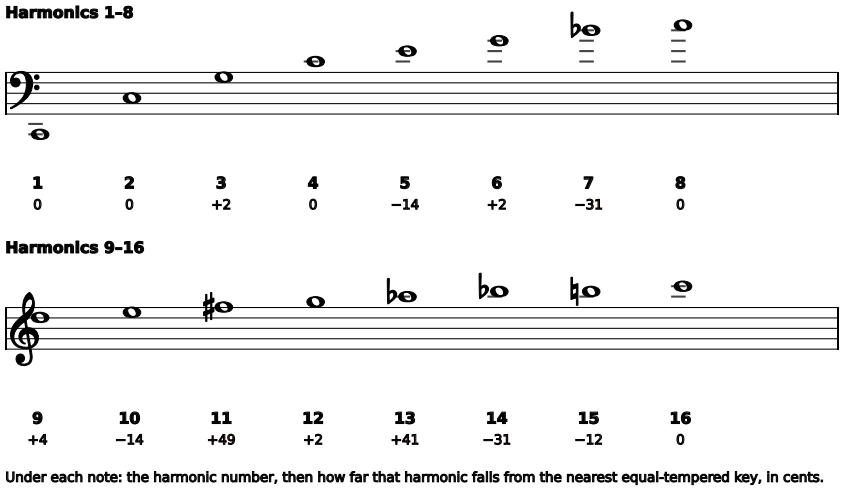


Figure 3.10: The harmonic series on C2, written out. The first few partials are the intervals a musician already knows: an octave, a fifth, another octave, a major third. Further up, the series starts to disagree with the piano, the seventh partial is a third of a semitone flat of any key, and the eleventh sits almost exactly between two. That disagreement is the subject of Chapter 9. Notice early that the notes in this figure are what a *string* produces, not what a keyboard can play.

Playing Harmonics

Touch a sounding string lightly at its midpoint. You have forced a node there. Every mode with a node at the midpoint: the 2nd, 4th, 6th, …, survives untouched; every mode without one: the 1st, 3rd, 5th, …, is killed. What remains sounds at $2f_1$, an octave above.

Touch at a third of the length and the surviving modes are the 3rd, 6th, 9th, …, sounding at $3f_1$, an octave and a fifth above. Touch at a quarter and you get $4f_1$, two octaves up.

This is why the technique is called "playing harmonics", and it is a direct, audible demonstration that the modes really are all present at once in an ordinary plucked note. You are not adding anything by touching the string. You are subtracting.

3.5 Standing Waves in an Air Column

Open and Stopped Pipes

The same argument applies to the air in a tube, with different boundary conditions.

At a **closed** end the air cannot move, so there is a displacement node. At an **open** end the air is free to move and the pressure is forced to atmospheric, so there is a displacement antinode.

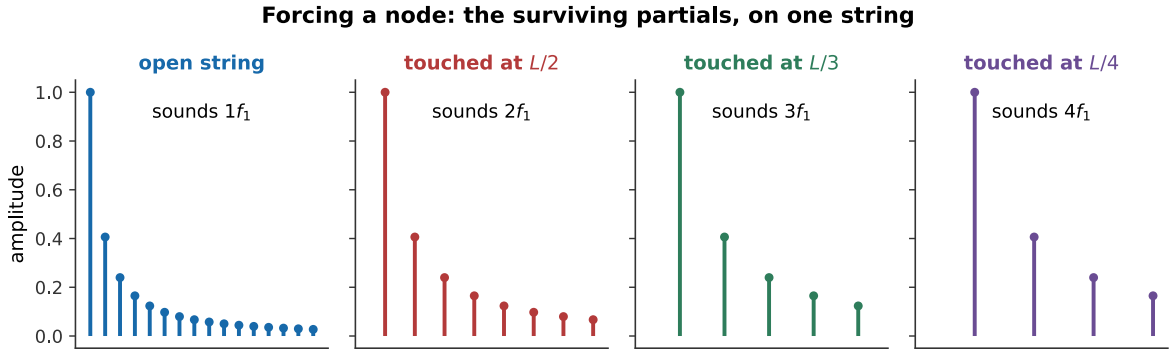


Figure 3.11: One string, touched at four places. The pitch rises through the harmonic series, and the tone becomes progressively thinner, because each touch removes more partials than the last. The glassy quality of a string harmonic is the sound of a spectrum with most of its members missing. *What you hear:*Four notes on one string. The first is the full open string; the next three rise by an octave, an octave and a fifth, and two octaves, and each sounds purer and glassier than the last. Listen: String Harmonics

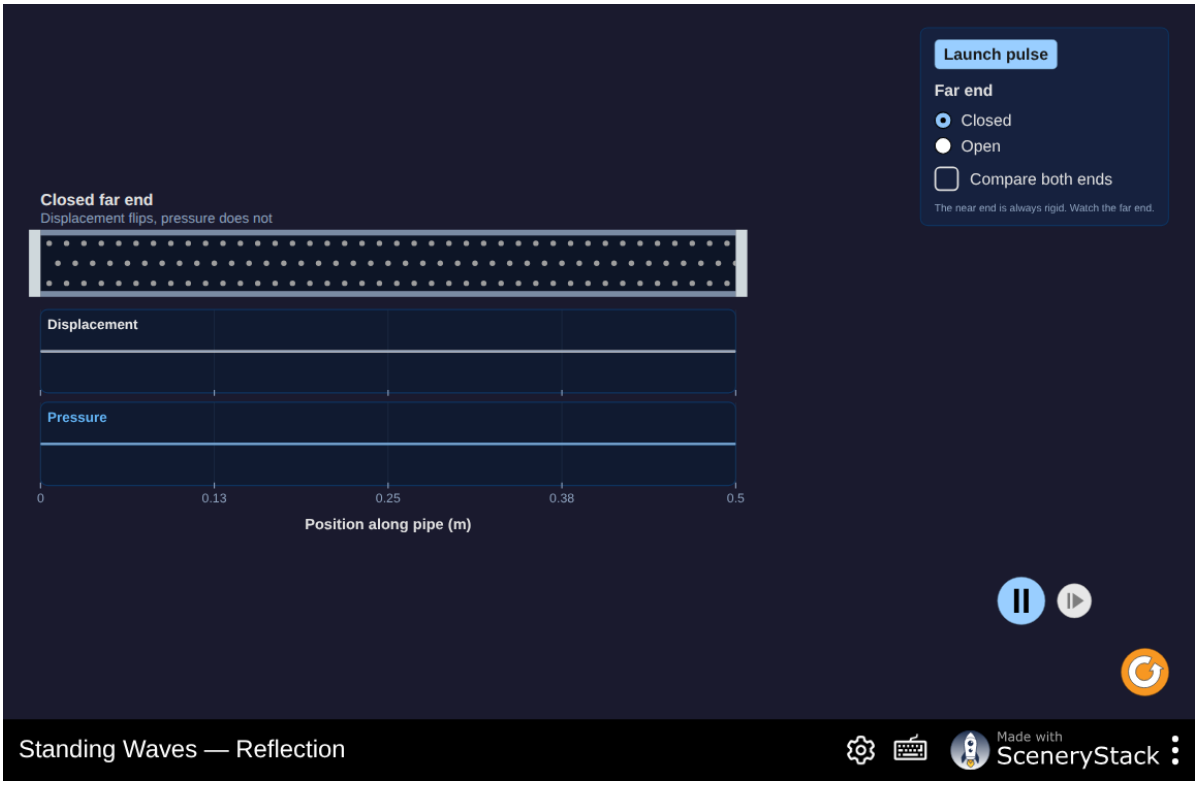


Figure 3.12: Drive one end of a string and sweep the frequency slowly. Almost everywhere nothing much happens; at a handful of frequencies the string leaps into a clean pattern. Count the loops and check against $f_n = nf_1$, then change the tension and watch every resonance move by the same factor. Interactive simulation: Standing Waves



Figure 3.13: Successive resonances on a driven string make the nodes, antinodes, and first several harmonics directly visible. Each higher mode fits one additional half-wavelength between the fixed ends. Watch on YouTube: Standing Wave Harmonics

A pipe **open at both ends** therefore needs antinodes at both ends. That is satisfied by the same family as the string, a whole number of half-wavelengths, and gives

$$f_n = n\,\frac{v}{2L}, \qquad n = 1, 2, 3, \ldots \tag{3.11}$$

A pipe **stopped at one end** needs a node at one end and an antinode at the other, which are a *quarter* wavelength apart. The shortest fit is $L = \lambda/4$, and the next fits are $3\lambda/4$, $5\lambda/4$, and so on:

$$f_n = n\,\frac{v}{4L}, \qquad n = 1, 3, 5, \ldots \tag{3.12}$$

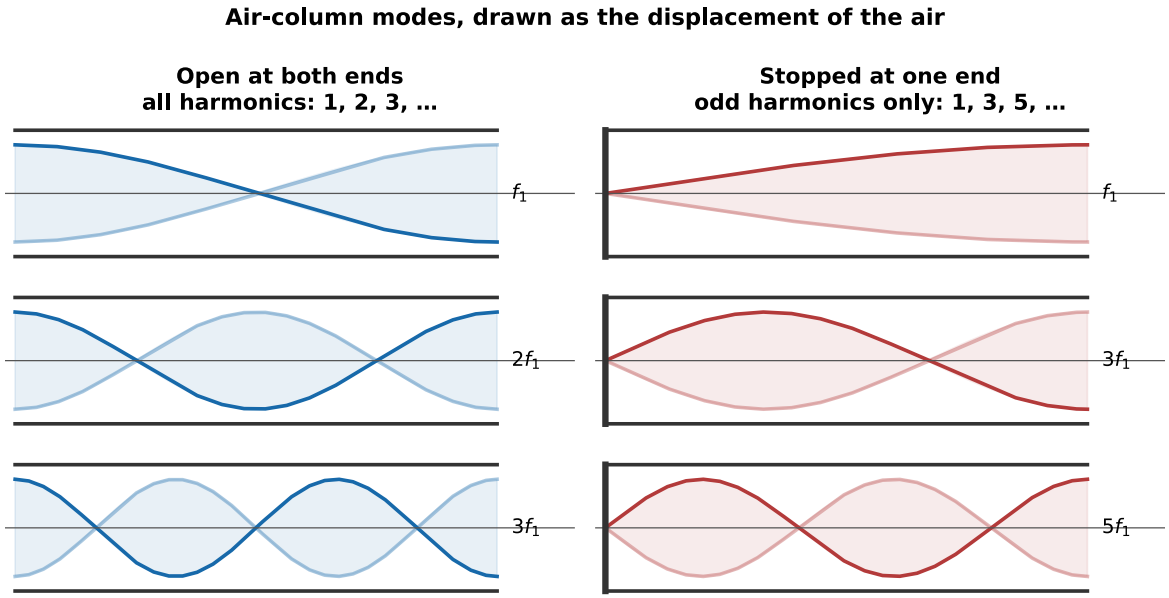


Figure 3.14: Air-column modes, drawn as the *displacement* of the air, oscillating inside their envelopes. The open pipe has antinodes at both ends and supports every harmonic. The stopped pipe has a node at the closed end and supports only the odd ones. Drawing the same modes as *pressure* would interchange nodes and antinodes everywhere, so a figure like this must say which quantity it shows.

Odd Harmonics and the Missing Even Ones

Two consequences follow from that pair of formulas, and both are audible.

A stopped pipe of a given length sounds an octave lower than an open pipe of the same length. Compare $v/4L$ with $v/2L$. Organ builders use this constantly: a stopped rank gives a 16-foot pitch from 8 feet of pipe, which is a considerable saving in a church.

A stopped pipe has only odd harmonics. This is not a small change of color. Removing every even partial removes the octave, the second octave, and the octave-plus-fifth from the spectrum, leaving a characteristic hollow, woody sound. It is why the clarinet, effectively a stopped pipe, since the reed end is closed, sounds so unlike the flute, which is an open one, and it is why a clarinet overblows to a twelfth rather than an octave (Chapter 11).

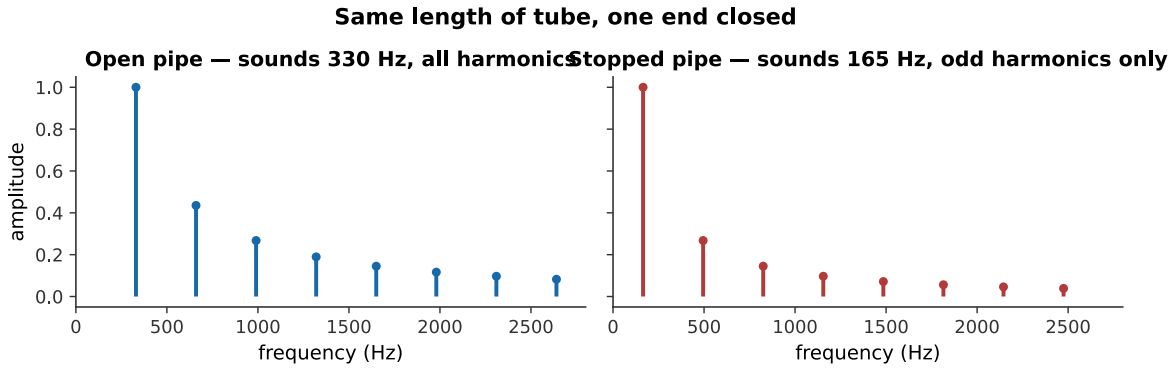


Figure 3.15: The same length of tube, played open and then stopped. Two things change at once; try to hear them separately: the pitch drops an octave, and the tone goes hollow because every even harmonic has vanished. *What you hear:* Two tones. The first is bright and full; the second is an octave lower and noticeably hollow, as though something had been taken out of the middle of it. Listen: Open pipe, Stopped pipe

Why the Pipe Length Is Not Quite the Wavelength

One correction, because it matters for anyone who tries to check these formulas by measurement.

The pressure at an open end is not *exactly* atmospheric, because the air just outside the tube is still being pushed around by the air inside it. The antinode therefore falls a little way *beyond* the physical opening, and the pipe behaves as though it were slightly longer than it is. The correction is about 0.6 times the tube's radius, per open end.

For a narrow pipe this is negligible. For a wide one it is not: a pipe 30 cm long and 5 cm across has an effective length of about 31.5 cm at one open end, or 33 cm at two: a pitch error of nearly a semitone if you ignore it. Chapter 11 treats the correction properly, and the laboratory exercise in Laboratory Exercises measures it.

> **Self-check**
> A stopped organ pipe sounds middle C, 262 Hz, at 20 °C. Without a calculator, say roughly how long it is. (Use $v \approx 343$ m/s, and remember which formula applies.)

3.6 Summary

- **Superposition**: where waves overlap, displacements add, and the waves emerge unchanged. This holds because the wave equation is linear, and it fails at extreme amplitudes, audibly so in a loud brass instrument and in a struck cymbal.
- **Interference** is superposition with a steady phase relationship. Two waves in step reinforce; two exactly opposed cancel. Energy is redistributed, not destroyed: the loud spots are four times the intensity of one source, not twice.
- **Path difference sets phase difference**: $\phi = (2\pi/\lambda)\Delta r$. Constructive interference requires $\Delta r = m\lambda$, destructive requires $\Delta r = (m + \frac{1}{2})\lambda$. The pattern is fine at high frequencies and coarse at low ones, which is why music through two speakers is colored rather than silenced.

- **Reflection inverts a pulse at a fixed end** and returns it upright at a free end. For air columns, a closed end is a displacement node and a pressure antinode; an open end is the reverse.
- **A standing wave is two identical waves traveling in opposite directions**, and it factorizes: $2A\sin(kx)\cos(\omega t)$. The shape is fixed in space and only its amplitude varies in time. Nodes never move and are half a wavelength apart.
- **A string fixed at both ends** admits only $\lambda_n = 2L/n$, giving $f_n = (n/2L)\sqrt{T/\mu}$: a complete harmonic series. Shorter, tighter, or lighter means higher, but tension enters as a *square root*, so an octave costs four times the tension.
- **Touching a string at a node** kills every mode that lacks one there, leaving a harmonic. This demonstrates that all the modes are present simultaneously in an ordinary note.
- **An open pipe gives $f_n = nv/2L$ with all harmonics; a stopped pipe gives $f_n = nv/4L$ with odd harmonics only.** A stopped pipe sounds an octave lower than an open pipe of the same length and sounds hollow, because every even partial is missing.
- **The effective length of a pipe exceeds its physical length** by about $0.6r$ per open end, which matters for wide tubes.

3.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity. The set mixes multiple choice, true/false, drag-the-words, fill-in-the-blank, and mark-the-words formats, and opens by counting nodes on a figure.

1. **Multiple choice, from a figure.** The activity reproduces the string-mode figure from earlier in this chapter. Its bottom panel shows the 4th harmonic of a string fixed at both ends. How many nodes lie strictly between the two fixed ends: two, three, four, or five?

2. **True or false.** Where two equal sources interfere constructively, the intensity is four times that of one source alone.

3. **Drag the words.** Complete: a pulse returns _____ from a fixed end and _____ from a free end; two loudspeakers reinforce where the path difference is a _____ number of wavelengths, and cancel where it includes one extra _____ wavelength. (Terms: inverted, upright, whole, half, doubled, quarter.)

4. **Fill in the blanks.** A pipe stopped at one end resonates at nv divided by _____ times L, sounds _____ harmonics only, and sits an _____ below an open pipe of the same length.

5. **Mark the words.** Identify the two changes that raise a string's fundamental frequency: "Shortening the string and tightening it both raise the pitch, while loosening it or fitting a heavier string lowers it."

3.8 Conceptual Questions

1. Two pulses of opposite sign meet on a string and momentarily cancel completely, leaving the string straight. Explain where the energy is at that instant.

2. Two loudspeakers play the same steady 1 kHz tone. A listener walks slowly across the room and hears the loudness rise and fall. Explain why playing music instead of a tone produces no such silences.

3. Explain why a pulse reflecting from a clamped end comes back inverted, using only the requirement that the clamped point cannot move.

4. A standing wave transports no energy along the string, yet the string is full of moving parts. Reconcile these two statements.

5. A guitarist touches a sounding string lightly at its exact midpoint and the pitch jumps up an octave. Explain what has been removed, and why touching at the midpoint removes it.

6. Explain why raising a string's tension by a factor of two does not raise its pitch by an octave, and state what factor would.

7. A clarinet and a flute of similar length sound roughly an octave apart. Explain which sounds lower and why, in terms of boundary conditions.

8. The end correction for an open pipe is roughly $0.6r$. Explain why this makes the correction important for an organ's largest pipes and negligible for a piccolo, even though the *fractional* pitch error is what matters.

3.9 Problems

Exercise 3.1
(Straightforward) Two loudspeakers, driven in step, play a 686 Hz tone. Take $v = 343$ m/s. (a) What is the wavelength? (b) What is the smallest non-zero path difference that produces silence at a listener? (c) The smallest that produces maximum loudness?

Solution to Exercise 3.1
(a) $\lambda = v/f = (343\text{ m/s})/(686\text{ Hz}) = 0.500$ m.
(b) Destructive interference needs $\Delta r = (m + \tfrac{1}{2})\lambda$; the smallest is

$$\Delta r = \tfrac{1}{2}(0.500\text{ m}) = 0.250\text{ m}. \tag{3.13}$$

(c) Constructive interference needs $\Delta r = m\lambda$. The smallest non-zero value is $\Delta r = 0.500$ m (the case $m = 0$, equal paths, is also loud but is not a *difference*).
Therefore, 25 cm of extra path silences the tone and 50 cm restores it.

Exercise 3.2
(Straightforward) A string 0.62 m long is fixed at both ends and waves travel along it at 240 m/s. Find the frequencies of its first four modes.

Solution to Exercise 3.2
The fundamental is

$$f_1 = \frac{v}{2L} = \frac{240\text{ m/s}}{2(0.62\text{ m})} = 194\text{ Hz}, \tag{3.14}$$

and the higher modes are whole-number multiples:

$$f_2 = 387\text{ Hz}, \qquad f_3 = 581\text{ Hz}, \qquad f_4 = 774\text{ Hz}. \tag{3.15}$$

Therefore, the first four modes are 194, 387, 581, and 774 Hz: an exact harmonic series.

Exercise 3.3
(Moderate) A guitar string of length 64.0 cm sounds 196 Hz when played open. (a) What is the wave speed along it? (b) Where must the player's finger go to sound 261.6 Hz? (c) What is the wavelength of the 196 Hz *sound* in air, and why is it so different from the wavelength on the string?

Solution to Exercise 3.3
(a) From $f_1 = v/2L$:

$$v = 2Lf_1 = 2(0.640\text{ m})(196\text{ Hz}) = 251\text{ m/s}. \tag{3.16}$$

(b) The tension and mass per length are unchanged, so v is unchanged and $f \propto 1/L$:

$$L' = L\,\frac{f_1}{f'} = (64.0\text{ cm})\frac{196}{261.6} = 47.9\text{ cm}, \tag{3.17}$$

measured from the bridge; so the finger goes $64.0 - 47.9 = 16.1$ cm from the nut.
(c) In air,

$$\lambda_{\text{air}} = \frac{343\text{ m/s}}{196\text{ Hz}} = 1.75\text{ m}, \tag{3.18}$$

against $2L = 1.28$ m on the string. They differ because the wave *speeds* differ, 251 m/s on the string against 343 m/s in air, while the frequency is necessarily shared, the string being what drives the air.

Exercise 3.4
(Straightforward) A string is tuned to 110 Hz. By what factor must its tension change to tune it to (a) 165 Hz, (b) 220 Hz? (c) Comment on whether (b) is practical.

Solution to Exercise 3.4
Since $f \propto \sqrt{T}$, the tension ratio is the square of the frequency ratio.
(a) $(165/110)^2 = 1.5^2 = 2.25$.
(b) $(220/110)^2 = 2^2 = 4.00$.
(c) Quadrupling the tension is not practical. String tensions are already set close to the breaking point of the material for good reasons, higher tension gives a louder, brighter note, so a factor of four would break the string, and if it did not, it would very likely pull the instrument apart. This is exactly why instruments span their range by changing length and mass per unit length rather than tension alone.

Exercise 3.5
(Straightforward) An organ pipe open at both ends is 1.10 m long. Take $v = 343$ m/s. (a) What is its fundamental frequency? (b) What are its next two modes? (c) If one end is stopped, what does the fundamental become?

Solution to Exercise 3.5
(a) For an open pipe, $f_1 = v/2L$:

$$f_1 = \frac{343\text{ m/s}}{2(1.10\text{ m})} = 156\text{ Hz}. \tag{3.19}$$

(b) All harmonics are present: $f_2 = 312$ Hz, $f_3 = 468$ Hz.
(c) For a stopped pipe, $f_1 = v/4L$:

$$f_1 = \frac{343 \text{ m/s}}{4(1.10 \text{ m})} = 77.9 \text{ Hz}. \tag{3.20}$$

Therefore, stopping one end drops the pitch by exactly an octave, and the surviving modes become 77.9, 234, 390 Hz: the odd harmonics only.

Exercise 3.6
(Moderate) A stopped organ pipe is to sound 131 Hz at 20 °C. (a) What length of pipe is needed, ignoring the end correction? (b) If the pipe is 4.0 cm in radius, what is the end correction, and what length should actually be cut?

Solution to Exercise 3.6
(a) From $f_1 = v/4L$:

$$L = \frac{v}{4f_1} = \frac{343 \text{ m/s}}{4(131 \text{ Hz})} = 0.655 \text{ m}. \tag{3.21}$$

(b) A stopped pipe has one open end, so one correction of about $0.6r$:

$$\Delta L = 0.6(0.040 \text{ m}) = 0.024 \text{ m}. \tag{3.22}$$

The *effective* length must be 0.655 m, so the physical length cut should be

$$0.655 - 0.024 = 0.631 \text{ m}. \tag{3.23}$$

Therefore, cutting the naive 65.5 cm would give an effective length of 67.9 cm and a pitch of 126 Hz, about 67 cents flat, two-thirds of a semitone, and grossly out of tune.

Exercise 3.7
(Moderate) A violin string sounds 440 Hz. A player touches it lightly one-third of the way along. (a) What frequency sounds? (b) Which harmonics survive? (c) Why does the note sound thinner than the open string?

Solution to Exercise 3.7
(a) Touching at $L/3$ forces a node there, which only modes with n a multiple of 3 possess. The lowest survivor is $n = 3$:

$$f = 3 \times 440 \text{ Hz} = 1320 \text{ Hz}. \tag{3.24}$$

(b) Harmonics 3, 6, 9, 12, …of the original string: that is, 1, 2, 3, 4, …of the new sounding pitch.

(c) Because two-thirds of the original partials have been removed. The surviving set is still a complete harmonic series on 1320 Hz, but each of its members was a *high* partial of the original string and was therefore weak to begin with. The result is a spectrum with far less energy in it and fewer strong partials, which the ear reports as thin and glassy.

Exercise 3.8
(Moderate) Two identical waves, each of amplitude A, interfere. (a) What is the resulting amplitude when they are exactly in step? (b) When exactly opposed? (c) By what factor is the intensity of the in-step case greater than that of one wave alone, and where does the extra energy come from?

Solution to Exercise 3.8

(a) $2A$. (b) Zero.

(c) Intensity goes as the square of amplitude, so the in-step case has $(2A)^2/A^2 = 4$ times the intensity of one wave alone, not twice.

The extra energy comes from the places where the waves cancel. Averaged over all positions, the mean intensity is $2A^2$, exactly twice one wave's, as conservation of energy requires. Interference redistributes energy from the nulls into the maxima; it does not create it.

Exercise 3.9

(Straightforward) The A string of a bass is 1.06 m long, of mass per unit length $\mu = 1.5 \times 10^{-2}$ kg/m, and sounds 55 Hz. (a) What tension is it under? (b) The string is replaced by one of twice the mass per unit length, tuned to the same pitch. What tension is now needed?

Solution to Exercise 3.9

(a) From $f_1 = (1/2L)\sqrt{T/\mu}$:

$$T = \mu\,(2Lf_1)^2 = (1.5 \times 10^{-2})\,[2(1.06)(55)]^2 = (1.5 \times 10^{-2})(116.6)^2 = 204 \text{ N}. \tag{3.25}$$

(b) At fixed L and f_1, $T \propto \mu$, so doubling the mass per unit length doubles the required tension:

$$T' = 408 \text{ N}. \tag{3.26}$$

Therefore, the original tension is about 204 N, roughly the weight of a 21 kg mass, and the heavier string would need twice that. This is the trade-off behind wound strings: mass lowers the pitch, but keeping the pitch while adding mass costs tension.

Exercise 3.10

(Challenging) Two speakers 3.0 m apart face a listener standing 8.0 m from the midpoint between them, on the perpendicular bisector. (a) What is the path difference? (b) The listener now walks 1.0 m to one side. Find the new path difference, and state whether a 1720 Hz tone is loud or quiet there. Take $v = 343$ m/s.

Solution to Exercise 3.10

(a) On the perpendicular bisector the two distances are equal, so $\Delta r = 0$ and the listener is at a maximum.

(b) Put the speakers at $y = \pm 1.5$ m and the listener at $(8.0, 1.0)$. Then

$$r_1 = \sqrt{8.0^2 + (1.0 - 1.5)^2} = \sqrt{64.0 + 0.25} = 8.0156 \text{ m}, \tag{3.27}$$

$$r_2 = \sqrt{8.0^2 + (1.0 + 1.5)^2} = \sqrt{64.0 + 6.25} = 8.3815 \text{ m}, \tag{3.28}$$

so $\Delta r = 0.366$ m. The wavelength is $\lambda = (343 \text{ m/s})/(1720 \text{ Hz}) = 0.1994$ m, so

$$\frac{\Delta r}{\lambda} = \frac{0.366}{0.1994} = 1.84. \tag{3.29}$$

Therefore, the path difference is 1.84 wavelengths, close to 1.5 or 2.5? It is nearer the whole number 2 than the half-integer 1.5, so the listener is closer to a loud position than a quiet one, though not exactly at a maximum.

Exercise 3.11
(Challenging) Show that the nodes of the standing wave $y = 2A\sin(kx)\cos(\omega t)$ are spaced half a wavelength apart, and that the antinodes lie exactly midway between them.

Solution to Exercise 3.11
A node is a position where $y = 0$ for all t. Since $\cos(\omega t)$ is not always zero, this requires

$$\sin(kx) = 0 \;\Rightarrow\; kx = m\pi \;\Rightarrow\; x = \frac{m\pi}{k} = \frac{m\lambda}{2}, \tag{3.30}$$

using $k = 2\pi/\lambda$. Successive nodes therefore differ by $\lambda/2$.

An antinode is a position where $|y|$ reaches its largest value, which requires $|\sin(kx)| = 1$, that is

$$kx = \left(m + \tfrac{1}{2}\right)\pi \;\Rightarrow\; x = \frac{(2m+1)\lambda}{4}. \tag{3.31}$$

These are the odd multiples of $\lambda/4$, which lie exactly halfway between the even multiples that give the nodes.

Therefore, nodes and antinodes alternate at intervals of a quarter wavelength, with each kind spaced half a wavelength from its own neighbors.

Exercise 3.12
(Moderate) A clarinet behaves as a pipe stopped at the reed end. Its lowest note is written E_3, sounding about 147 Hz. (a) Estimate its sounding length at 20 °C. (b) A flute of the same length behaves as an open pipe; what is its lowest note? (c) Explain why the clarinet is the longer-sounding instrument for its physical size.

Solution to Exercise 3.12
(a) For a stopped pipe, $L = v/4f_1$:

$$L = \frac{343\ \text{m/s}}{4(147\ \text{Hz})} = 0.583\ \text{m}. \tag{3.32}$$

(b) For an open pipe of the same length, $f_1 = v/2L = (343\ \text{m/s})/(2 \times 0.583\ \text{m}) = 294$ Hz, exactly an octave higher.

(c) Because a stopped pipe sounds at $v/4L$ while an open one sounds at $v/2L$. The stopped pipe gets an octave lower out of the same physical length, since the shortest standing wave it can hold is a quarter wavelength rather than a half. A clarinet is therefore about half the length of a flute sounding the same low note: a real and substantial advantage, paid for by having only odd harmonics.

Exercise 3.13
(Moderate) A stopped pipe and an open pipe are both tuned to sound 220 Hz. (a) Find the length of each. (b) List the first four modes of each. (c) A listener hears both play their second mode. What interval do they hear?

Solution to Exercise 3.13

(a) Stopped: $L = v/4f_1 = (343\text{ m/s})/(4 \times 220\text{ Hz}) = 0.390$ m. Open: $L = v/2f_1 = (343\text{ m/s})/(440\text{ Hz}) = 0.780$ m.

(b) Stopped, odd harmonics only: 220, 660, 1100, 1540 Hz. Open, all harmonics: 220, 440, 660, 880 Hz.

(c) The stopped pipe's second mode is 660 Hz; the open pipe's is 440 Hz. The ratio is $660/440 = 3/2$.

Therefore, the listener hears a perfect fifth, and one which is *exactly* 3:2, since both frequencies are exact harmonics of the same fundamental. Chapter 9 takes up what happens when a keyboard has to approximate that ratio.

Exercise 3.14

(Moderate) A pipe open at both ends is 28.0 cm long and 3.0 cm in radius. (a) Find its fundamental ignoring end corrections. (b) Find it including them. (c) Express the difference in cents, and say whether it matters.

Solution to Exercise 3.14

(a) Ignoring corrections:

$$f_1 = \frac{v}{2L} = \frac{343\text{ m/s}}{2(0.280\text{ m})} = 613\text{ Hz}. \tag{3.33}$$

(b) Two open ends, each adding about $0.6r = 0.018$ m:

$$L_{\text{eff}} = 0.280\text{ m} + 2(0.018\text{ m}) = 0.316\text{ m}, \qquad f_1 = \frac{343\text{ m/s}}{2(0.316\text{ m})} = 543\text{ Hz}. \tag{3.34}$$

(c) In cents:

$$n = 1200\log_2\left(\frac{543}{613}\right) = 1200(-0.1750) = -210\text{ cents}. \tag{3.35}$$

Therefore, ignoring the correction would put the pipe more than two semitones sharp of where it actually sounds. For a tube this wide relative to its length, the end correction is not a refinement, it dominates any attempt to predict the pitch.

Chapter 4

Resonance and Normal Modes

Learning Objectives

By the end of this chapter, you should be able to:

- Distinguish free vibration at a natural frequency from forced vibration at a driving frequency.
- Sketch the response of a driven oscillator against driving frequency, and identify the resonance peak, its height, and its width.
- Explain why a system driven at resonance builds up a large amplitude, in terms of energy delivered in step with the motion.
- Relate damping to the sharpness of a resonance, define the quality factor Q, and calculate it from the bandwidth of a measured response curve.
- Explain the trade-off a Q value represents for an instrument: a sharp resonance is loud and selective, a broad one is even and responsive.
- Define a normal mode, and explain why a system with N degrees of freedom has N of them.
- Explain how an arbitrary motion of a vibrating system can be written as a combination of its normal modes, and why the modes are the natural description.
- Identify the resonator in a guitar, a violin, a trumpet, and the human voice, and describe what each contributes to the sound.

Introduction

Chapter 3 established that a string fixed at both ends can only vibrate at certain frequencies. That is a statement about what the string will do if left alone. This chapter is about what happens when it is *not* left alone, when something drives it.

The answer is the single most useful idea in instrument design. Drive a system at one of its natural frequencies and it responds enormously; drive it anywhere else and it barely responds at all. That selectivity is **resonance**, and every instrument in this book depends on it twice over: once to choose which frequency will sound, and again to get that frequency out into the room loudly enough to hear.

Two ideas run through the chapter, and they turn out to be the same idea seen from two sides.

The first is the **trade-off**. A sharply tuned resonator responds hugely, but only to a narrow band of frequencies, and it rings for a long time after the driving stops. A broadly tuned one

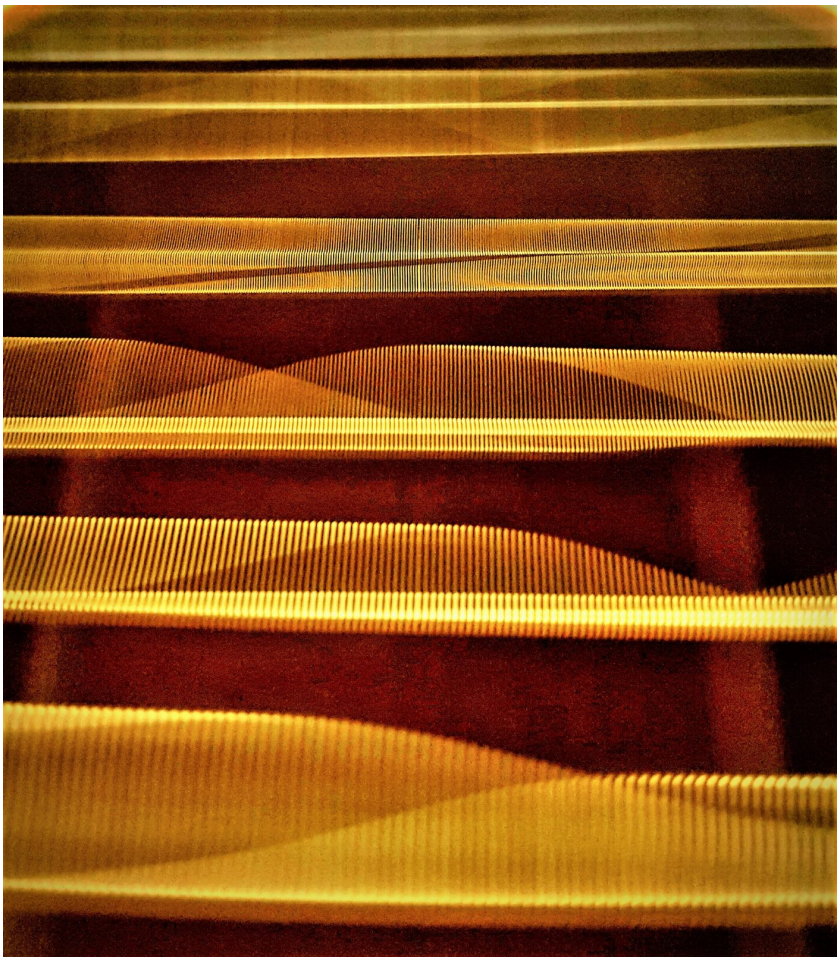


Figure 4.1: A guitar is a compact collection of resonators: strings, plates, an enclosed air cavity, and the room beyond them. Here the strings themselves are doing it, each blurred band the shape of one string's vibration. Matt Stetzel, CC BY 4.0.

responds modestly, to almost anything, and stops as soon as the driving does. An instrument maker must choose, and different instruments choose differently.

The second is the **normal mode**. A real object does not have one natural frequency; it has many, each with its own characteristic shape. The remarkable fact, and the reason Chapter 5 can exist, is that any motion whatsoever of such an object is a sum of its modes.

4.1 Free Vibration and Natural Frequency

Natural Frequencies of Real Objects

Displace something and let go, and it vibrates at its **natural frequency**. A mass on a spring does it at $(1/2\pi)\sqrt{k/m}$; a string fixed at both ends does it at $(1/2L)\sqrt{T/\mu}$; a wine glass, a tuning fork, a wooden bar, a column of air, a bridge, and a building all do it at frequencies set by their own stiffness and mass.

This is **free vibration**: no one is driving it, the amplitude decays as Chapter 1 described, and the frequency is whatever the object's construction dictates.

Notice that damping barely changes the frequency. A lightly damped oscillator rings at almost exactly the frequency it would have had with no damping at all: the correction is of order $1/Q^2$, and for any Q above about 5 it is negligible. A guitar string therefore sounds the same note whether you let it ring or mute it quickly, and the formulas of Chapter 3, which ignored damping entirely, were safe to write down.

Why Every Object Has More Than One

A mass on a spring has exactly one natural frequency, because it has exactly one way to move: the mass goes left or it goes right.

A string is different. It can bow out in one arch, or in two, or in three, and each of those shapes has its own frequency. The number of natural frequencies is the number of independent ways the object can move, its number of **degrees of freedom**, and for anything continuous, that number is unlimited.

This matters for a reason that will not be obvious yet: it means the word "resonance" is almost always plural. A guitar body does not have *a* resonance; it has dozens, and the shape of the resulting response curve is what a luthier is actually adjusting when they thin a brace by a fraction of a millimeter.

4.2 Driven Oscillation and Resonance

Driving an Oscillator Off and On Resonance

Now apply a force that oscillates at a frequency f of your choosing, and ask how large a response you get.

After any initial transient dies away, the system settles into oscillating **at the driving frequency**, not at its own. That is worth stating clearly because it is counterintuitive: the driven system abandons its natural frequency and does what it is told. What its natural frequency controls is not the frequency of the response but its **size**.

- Drive it very slowly, far below f_0: the mass simply follows the force. The response is small and in step with the drive.

- Drive it at f_0: each push arrives exactly when it helps, cycle after cycle, and the amplitude builds far beyond anything the force could achieve on its own.
- Drive it far above f_0: the mass cannot keep up. It barely moves, and what motion there is lags half a cycle behind.

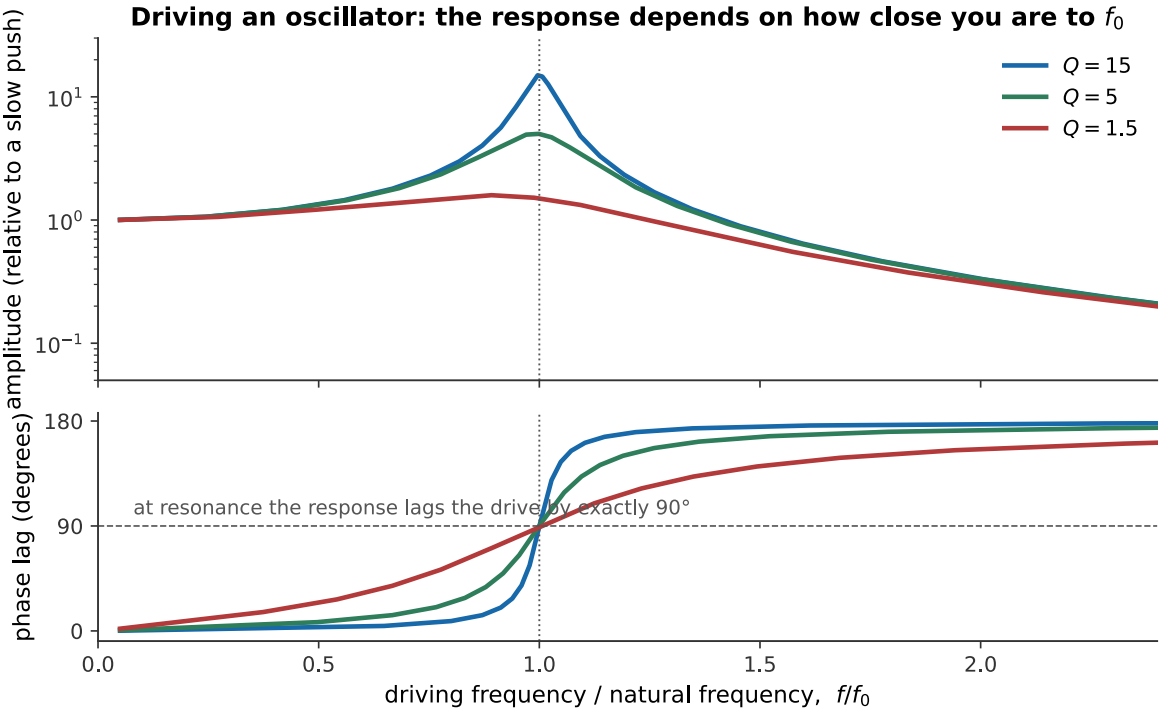


Figure 4.2: The response of a driven oscillator as the drive frequency sweeps back and forth through resonance. **Left**: a constant-strength force drives the $Q = 5$ oscillator shown by the green curve. Its motion grows dramatically near $f = f_0$, and its timing shifts from following the force to opposing it. **Right, top**: the amplitude for three amounts of damping. **Right, bottom**: the corresponding phase lag. At resonance the response lags the drive by exactly 90°, whatever the damping, which is the precise statement of "pushing at the right moment".

The Response Curve

The curve in that figure is the object an experimenter measures; reading it is a skill of its own.

Its **peak** is at the natural frequency. Its **height** says how much amplification the resonance provides. Its **width** says how fussy the resonator is about frequency. Height and width are not independent: a taller peak is always a narrower one, for reasons the next section makes precise.

> **Note**
> The most famous demonstration of resonance, a singer shattering a wine glass, is real but demanding. It needs the singer to hit the glass's natural frequency within a few hertz, to hold it, and to be loud enough that a Q of several hundred multiplies the amplitude past the glass's breaking strain. Amplification is usually involved. The physics is exactly the curve above; what makes it hard is the narrowness of the peak.

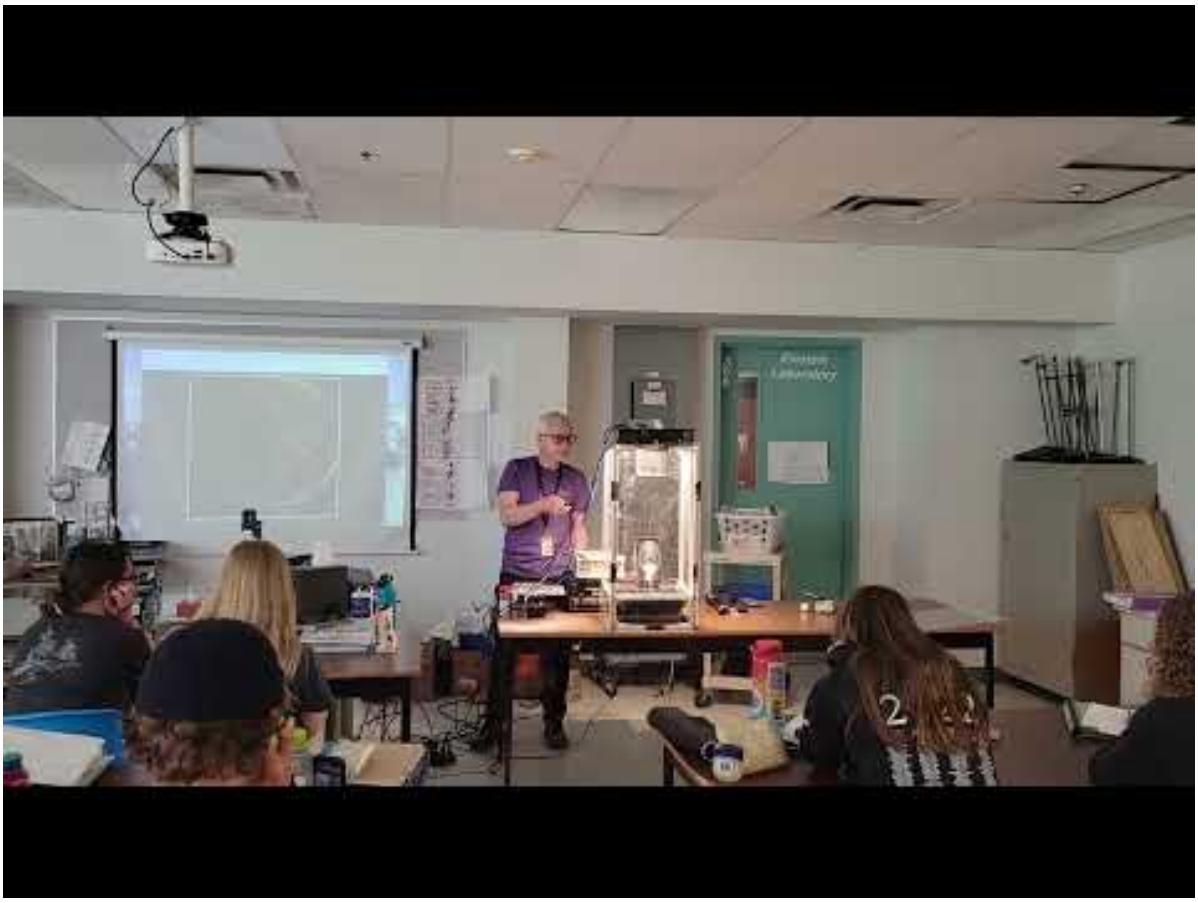


Figure 4.3: In this University of British Columbia outreach demonstration, a loudspeaker drives a wine glass at its natural frequency. High-speed footage shows the resonant motion growing until the glass fails: a vivid example of a small periodic force transferring energy efficiently when its frequency is correctly matched. Watch on YouTube: Resonance: Breaking a Wine Glass

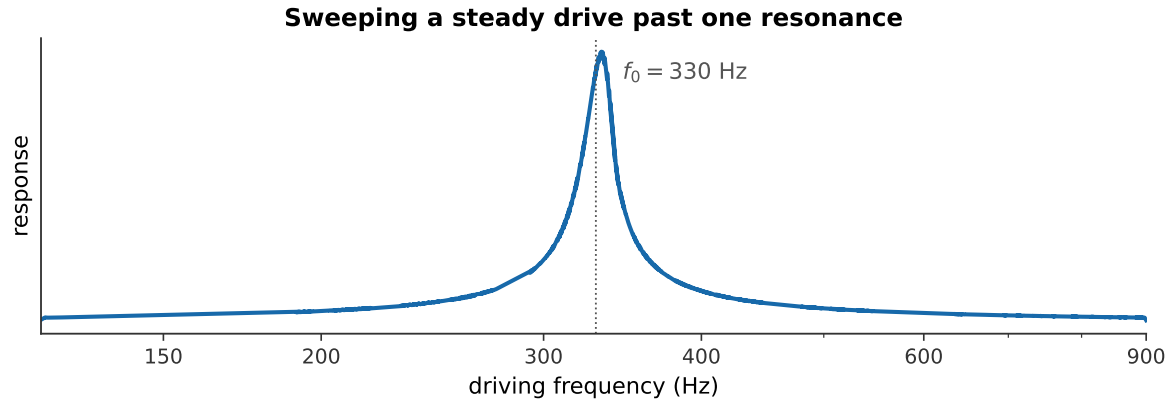


Figure 4.4: A steady drive of constant strength, swept slowly from 120 Hz to 900 Hz past a resonance at 330 Hz. The *drive* does not change in strength at any point. Everything you hear is the resonator's selectivity. *What you hear:*A tone gliding slowly upward. For most of the sweep it is barely audible; as it passes about 330 Hz it swells dramatically, then dies away again as the glide continues. Listen: Resonance Sweep

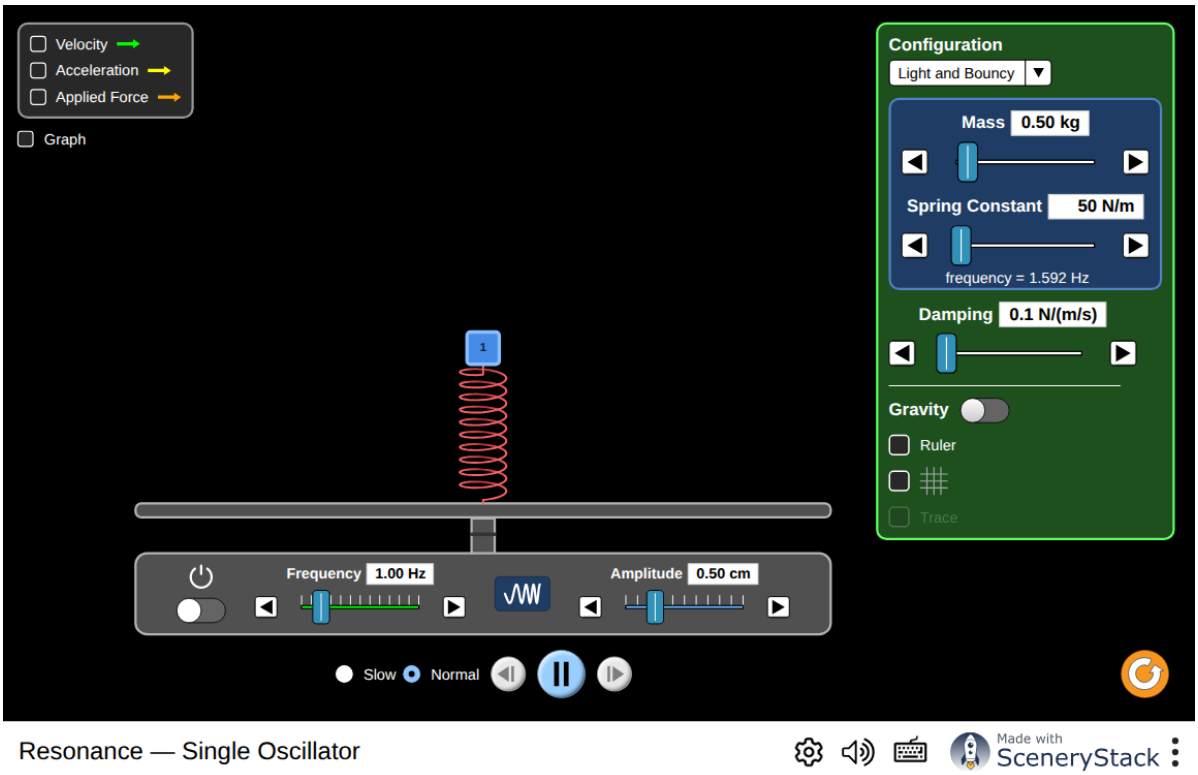


Figure 4.5: Set the damping low and sweep the driving frequency slowly through the natural frequency. The mass always moves at the *driving* frequency; it is the amplitude that changes, rising sharply at resonance. Increase the damping and repeat: the peak becomes lower and broader. Then change the mass or spring constant, predict which way the resonance will move, and test your prediction. Interactive simulation: Resonance — single oscillator

Phase: Pushing at the Right Moment

Why does driving at f_0 work so well? The answer is about timing, and it is better than the usual "the pushes add up".

To feed energy into an oscillator you must push in the direction it is already moving, push in phase with the **velocity**, not with the displacement. For a sinusoid, velocity leads displacement by a quarter cycle. So the drive must lead the displacement by 90°, which is exactly the phase relationship the bottom panel of the figure shows at resonance.

Off resonance, the drive and the motion drift in and out of step, so some pushes add energy and others take it away, and the amplitude never builds. At resonance the relationship is locked, and every push helps.

This is also why a child on a swing must push at the right moment rather than merely often, and why pushing twice as often does not work at all.

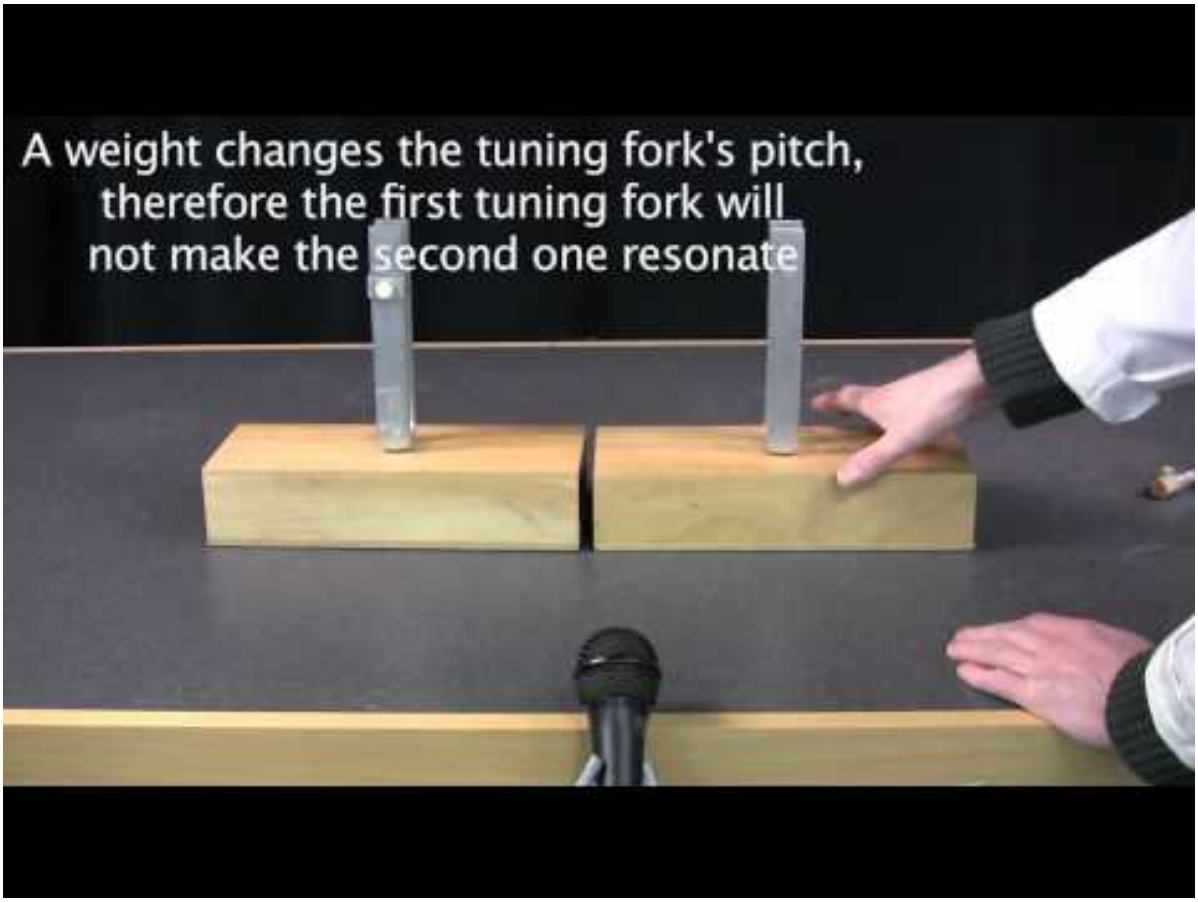


Figure 4.6: MIT's Physics Instructional Resources Lab demonstrates natural frequency and sympathetic vibration with matched tuning forks. One resonator responds strongly because the other drives it at the frequency it already prefers. Watch on YouTube: Tuning Forks

4.3 Damping, Bandwidth, and the Quality Factor

Damping Sets the Height and the Width

What stops the amplitude at resonance from growing without limit is damping. Energy is fed in by the drive and lost to friction and radiation, and the amplitude settles where the two balance.

So a lightly damped system reaches a very large amplitude before it balances: a tall peak. It is also fussy: because it loses so little energy per cycle, it takes many cycles to build up, and during those cycles the drive must stay in step, which it will only do if the frequency is very close to f_0. Tall and narrow go together.

A heavily damped system balances quickly at a modest amplitude and is not fussy, because it never needed many consecutive well-timed pushes. Short and broad go together.

Defining Q

The **quality factor** Q measures this, and it can be defined in three equivalent ways:

$$Q = \frac{f_0}{\Delta f} = 2\pi \frac{\text{energy stored}}{\text{energy lost per cycle}} = \pi f_0 \tau, \tag{4.1}$$

where Δf is the width of the response peak between the points at which the amplitude has fallen to $1/\sqrt{2}$ of its maximum, and τ is the free-decay time of Chapter 1.

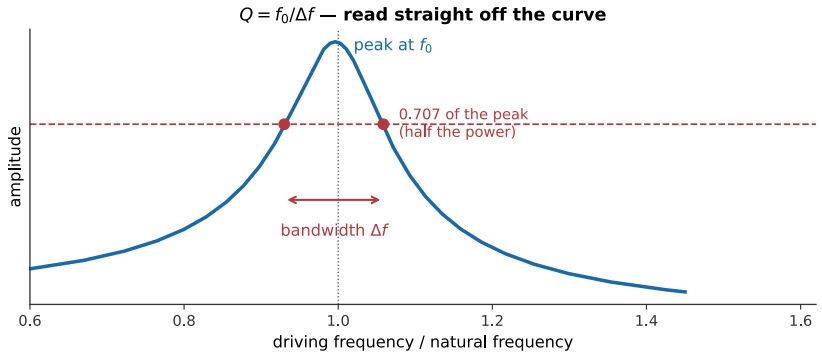


Figure 4.7: How Q is measured. Find the peak, drop to 0.707 of its height, which is half the *power*, since power goes as amplitude squared, and measure the width between the two crossings. The quality factor is the peak frequency divided by that width.

The third definition, $Q = \pi f_0 \tau$, is the one that connects this chapter to the last. It says that **a narrow resonance and a long ring are the same fact**, measured two different ways. Something that takes a long time to stop is something that is fussy about what starts it.

The Shape of the Response Curve

Q summarizes the response curve with a single number, its ratio of peak frequency to width. The curve itself, amplitude against driving frequency, has a definite shape, and it is worth writing down, because it is what a real measurement produces and what the rest of the book quietly assumes.

Write the driving frequency as a ratio to the natural frequency, $r = f/f_0$, so $r = 1$ is exactly on resonance. The steady-state amplitude, relative to how far the same force would push the system if applied very slowly ($r \to 0$), is

$$\frac{A}{A_{\text{static}}} = \frac{1}{\sqrt{(1 - r^2)^2 + (r/Q)^2}}. \tag{4.2}$$

Three checks confirm this is the curve already described in words. Far below resonance ($r \ll 1$), both terms under the root vanish and $A/A_{\text{static}} \to 1$: the mass simply follows the force, as §4.2 said. Exactly on resonance ($r = 1$), the first term vanishes and only the damping term survives, leaving $A/A_{\text{static}} = Q$: the peak height is nothing but Q itself, restated as a formula. Far above resonance ($r \gg 1$), the $(1 - r^2)^2$ term dominates and the response falls as $1/r^2$: the mass cannot keep up, and doubling the driving frequency above resonance cuts the response to a quarter.

Where the response formula comes from

The equation of motion for a mass m on a spring k with damping constant b, driven by a force $F_0 \cos(2\pi f t)$, is

$$m\ddot{x} + b\dot{x} + kx = F_0 \cos(2\pi f t). \tag{4.3}$$

After the initial transient dies away, the system settles into oscillating at the driving frequency, $x(t) = A\cos(2\pi f t - \phi)$, for some amplitude A and phase lag ϕ. Substituting this trial solution and collecting the sine and cosine parts separately gives two equations, which combine to

$$A = \frac{F_0}{\sqrt{(k - m\omega^2)^2 + (b\omega)^2}}, \qquad \omega = 2\pi f. \tag{4.4}$$

The static response, at $\omega \to 0$, is just $A_{\text{static}} = F_0/k$. Dividing the general expression by this, and writing everything in terms of $r = f/f_0 = \omega/\omega_0$ and $Q = \sqrt{mk}/b$ (which is equivalent to the definitions already given), reduces the ratio to the compact form quoted above. None of the algebra changes the physics already argued from the shape of the curve; it only makes the peak height, the low- and high-frequency limits, and the $1/r^2$ falloff exact rather than qualitative.

Worked example: Q from a measurement

A guitar body's main air resonance is measured by driving it and recording the response. The peak is at 102 Hz, and the response falls to 0.707 of its peak at 94 Hz and 112 Hz. Find Q and the time the resonance would ring for.

The bandwidth is $\Delta f = 112 - 94 = 18$ Hz, so

$$Q = \frac{f_0}{\Delta f} = \frac{102 \text{ Hz}}{18 \text{ Hz}} = 5.7. \tag{4.5}$$

The decay time follows from $Q = \pi f_0 \tau$:

$$\tau = \frac{Q}{\pi f_0} = \frac{5.7}{\pi(102 \text{ Hz})} = 0.018 \text{ s}. \tag{4.6}$$

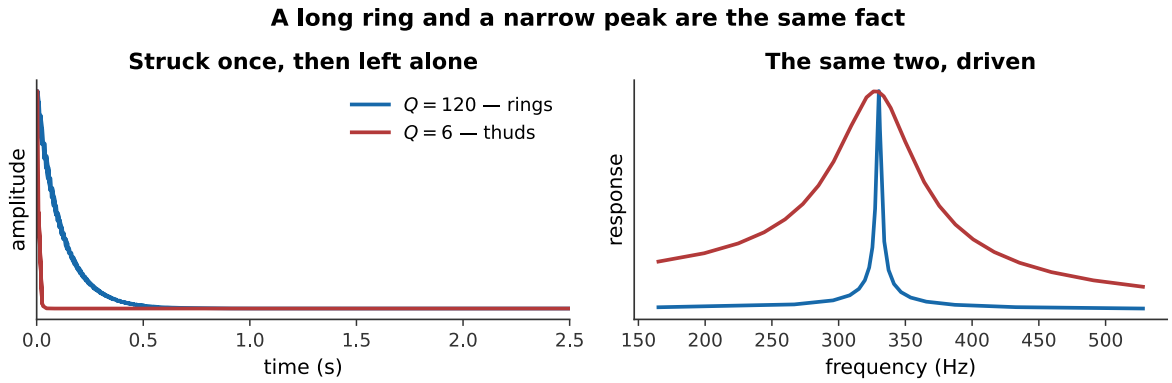


Figure 4.8: The same brief tap, through two resonators tuned to the same frequency and differing only in damping. The left panel of the figure shows what you hear; the right panel shows the response curves of the same two resonators when driven instead. A long ring and a narrow peak are two views of one property. *What you hear:*Two struck sounds at the same pitch. The first rings on clearly for a second or more, like a struck glass; the second is a dull thud that is over almost at once. Listen: Q = 120, Q = 6

That is a very low Q and a very short ring, under two hundredths of a second. It is exactly what a guitar needs: the body must respond to whatever note is played, not impose a note of its own, and it must follow the player's rhythm rather than smearing it.

The Instrument Maker's Trade-off

Now the trade-off can be stated properly, and it is the central design tension in acoustic instruments.

Energy radiated as sound is energy lost from the vibration, and energy lost from the vibration lowers Q. So:

$$\text{loud} \iff \text{low } Q \iff \text{short ring and broad response.} \tag{4.7}$$

	High Q	Low Q
Response	Huge, but only very near f_0	Modest, over a wide band
Ring	Long	Short
Loudness	Quiet	Loud
Examples	Tuning fork, wine glass, bell	Guitar body, violin body, drum head

A tuning fork has a Q of several thousand: it is a superb frequency standard and an almost useless loudspeaker, so it must be pressed against a table to be heard. A guitar body has a Q of around five: a poor frequency standard and an excellent radiator.

The requirement for a *resonator* in an instrument is therefore almost the opposite of the requirement for the *vibrator*. The string should have high Q, it holds the pitch and sustains the note. The body should have low Q, it should amplify everything the string offers, equally, and get it into the room. Chapter 10 shows how the bridge mediates between the two.

4.4 Normal Modes

Two Masses, Two Modes

Move from one mass to two, coupled by a spring between them, and something new appears.

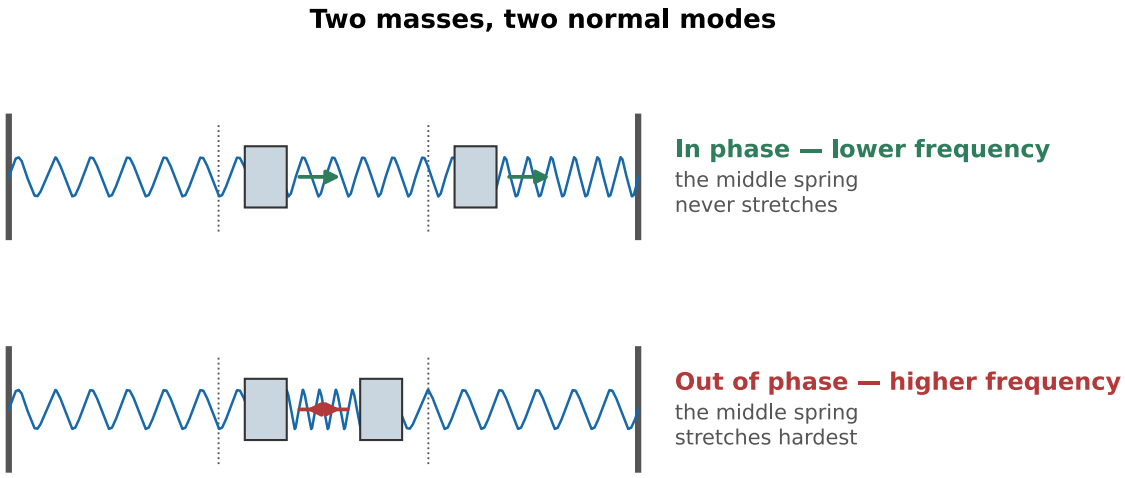


Figure 4.9: The two normal modes of a two-mass system. In the **in-phase** mode the masses move together and the connecting spring is never stretched, so it contributes no restoring force and the frequency is lower. In the **out-of-phase** mode they move oppositely, the connecting spring is stretched hardest, and the frequency is higher.

A **normal mode** is a pattern of motion in which every part of the system oscillates at the *same* frequency with a *fixed* relative amplitude and phase. Set a normal mode going and it persists, keeping its shape, exactly like a single mass on a single spring.

Two masses give two modes. The general rule is that a system with N degrees of freedom has exactly N normal modes, no more, and no fewer.

Start such a system in some arbitrary way, and it will *not* keep its shape: the motion looks complicated and never quite repeats. That is because you have excited both modes at once, and they are running at different frequencies.

Many Masses, Many Modes

Add masses and you add modes. A string is the limit of this process: infinitely many infinitesimal masses, and therefore infinitely many modes, which is exactly the family of standing waves Chapter 3 found.

The modes of a string are special in one respect, and the specialness is the reason music exists. Their frequencies are $f_n = nf_1$: exact whole-number multiples. Nothing guarantees this in general. The modes of a *drumhead* are in the ratios $1 : 1.59 : 2.14 : 2.30 : \ldots$, and the modes of a *bar* are in the ratios $1 : 2.76 : 5.40 : \ldots$: the reason Chapter 12 is a chapter about instruments with no definite pitch.

Any Motion Is a Sum of Modes

Here is the central result of the chapter.

Any motion of the system, however complicated, can be written as a sum of its normal modes, each with its own amplitude and phase. The modes form a complete set. Nothing the system can do lies outside them.

This transforms the problem of a plucked string from something apparently intractable, a triangle of string, released from rest, into bookkeeping. Decompose the initial shape into modes, let each mode oscillate at its own frequency, add them back up.

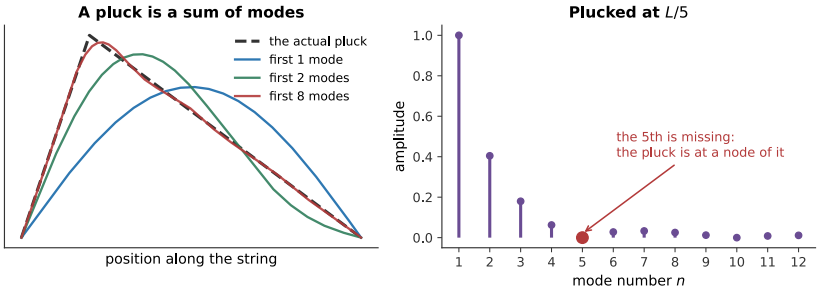


Figure 4.10: A string plucked one-fifth of the way along. **Left**: the triangular shape, rebuilt from its modes, one mode is a poor fit, eight are very nearly exact. **Right**: how much of each mode the pluck contains. The fifth mode is entirely absent, because the pluck point is a node of it: you cannot excite a mode by pulling on a place that mode does not move.

That missing fifth mode is not a curiosity; it is a technique. Guitarists pluck near the bridge for a bright sound and over the soundhole for a mellow one, and what they are doing is choosing which modes to leave out. Chapter 10 makes the rule explicit: **plucking at L/n silences the nth harmonic and every multiple of it.**

Finding the mode amplitudes: Fourier's method in advance

Decomposing the pluck is a calculation, and Chapter 5 is about to do the same thing to sound.

The modes of a string fixed at both ends are $\sin(n\pi x/L)$. The claim is that any initial shape $y(x)$ satisfying $y(0) = y(L) = 0$ can be written

$$y(x) = \sum_{n=1}^{\infty} c_n \sin\!\left(\frac{n\pi x}{L}\right). \tag{4.8}$$

To find a particular c_n, use the fact that the modes are **orthogonal**:

$$\int_0^L \sin\!\left(\frac{n\pi x}{L}\right)\sin\!\left(\frac{m\pi x}{L}\right)\mathrm{d}x = \begin{cases} L/2 & n = m \\ 0 & n \neq m. \end{cases} \tag{4.9}$$

Multiply both sides of the expansion by $\sin(m\pi x/L)$ and integrate over the string. Every term on the right vanishes except the one with $n = m$, leaving

$$c_m = \frac{2}{L}\int_0^L y(x)\,\sin\!\left(\frac{m\pi x}{L}\right)\mathrm{d}x. \tag{4.10}$$

For a string plucked to a height h at a point a along its length, this integral evaluates to

$$c_n = \frac{2hL^2}{\pi^2 n^2\, a(L-a)}\,\sin\!\left(\frac{n\pi a}{L}\right). \tag{4.11}$$

Two things fall out immediately. The $1/n^2$ makes the high modes weak, so a plucked string is not harsh. And the $\sin(n\pi a/L)$ vanishes whenever $a = L/n$: the missing-mode rule, now derived rather than asserted.

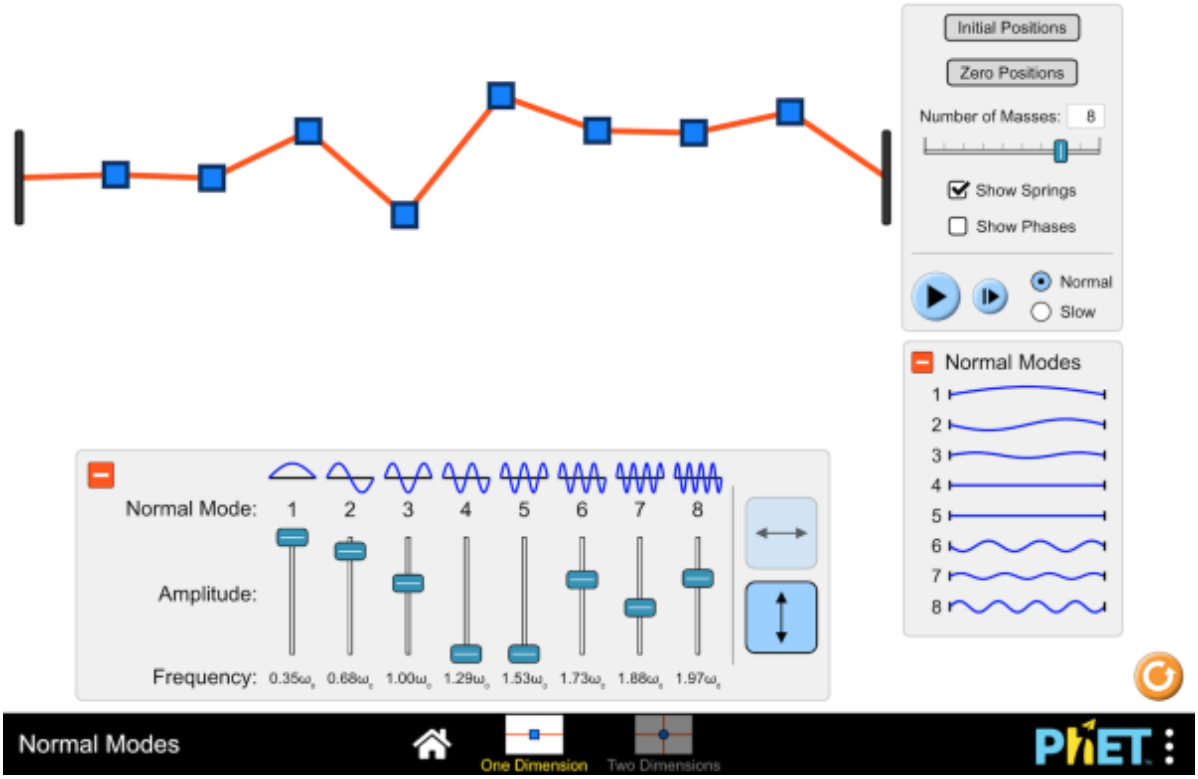


Figure 4.11: A chain of masses whose number you can change. Excite one mode at a time and watch it hold its shape; then excite two at once and watch the motion stop looking like anything in particular. Note that N masses always give exactly N modes. Interactive simulation: Normal Modes

4.5 Resonators in Musical Instruments

The Helmholtz Resonator

The simplest acoustic resonator is a cavity with a narrow opening: a bottle. Blow across the top and it sounds one note, and only one.

The mechanism is a mass on a spring, made entirely of air. The plug of air in the neck is the mass. The much larger body of air in the cavity is the spring: push the plug in and you compress the cavity, which pushes back. Its frequency is

$$f_0 = \frac{v}{2\pi}\sqrt{\frac{A}{V L_{\text{eff}}}}, \tag{4.12}$$

where A is the neck's cross-sectional area, L_{eff} its effective length, and V the cavity volume.

A guitar body is a Helmholtz resonator whose neck is the soundhole, and its air resonance, around 100 Hz on a typical steel-string, is deliberately placed near the bottom of the instrument's range, where the top plate alone radiates poorly. The same trick appears as a bass-reflex port in a loudspeaker cabinet (Chapter 15).

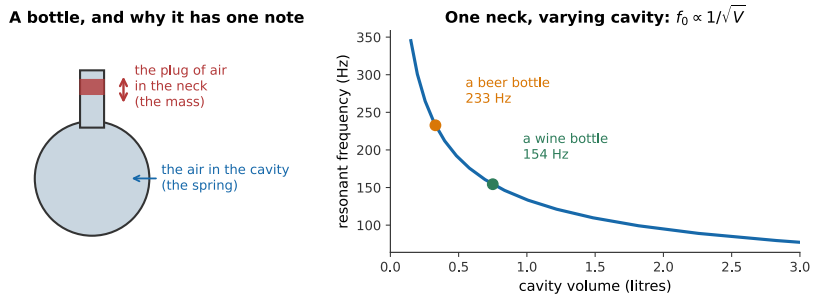


Figure 4.12: The Helmholtz resonator. Unlike a pipe, it has essentially **one** resonance, not a harmonic series, because there is no length along which standing waves can fit, only a single mass bouncing on a single spring. Filling the bottle with water reduces V and raises the pitch, in inverse proportion to the square root of what is left.

Bodies, Bores, and Cavities

Every acoustic instrument contains at least one resonator. The list below names which is which before Part V takes them one at a time.

Instrument	The vibrator	The resonator
Guitar, violin, piano	The string	The body, the soundboard, and the air inside
Flute, clarinet, trumpet	The air jet, reed, or lips	The air column in the bore
Timpani	The membrane	The kettle of air beneath it
Voice	The vocal folds	The vocal tract

The wind instruments are the odd family out, and the difference is fundamental. In a string instrument the resonator merely *responds* to the vibrator, which sets the pitch on its own. In a wind instrument the resonator **controls** the vibrator: the air column feeds back on the reed or the lips and tells them what frequency to oscillate at. That feedback is why a clarinetist can play a scale with one reed, and why Chapter 11 needs a different kind of argument from Chapter 10.

Resonance as Amplifier and as Filter

A final framing, which will be used repeatedly.

A resonator does not create energy. It is not an amplifier in the electrical sense, nothing is being added. What it does is provide a much better path for energy to leave the vibrator and enter the air, by presenting a large radiating surface where there was a thin string. The sound is louder and the note is shorter, and the total energy is unchanged.

A resonator is also a **filter**. It responds strongly to the frequencies near its peaks and weakly to the rest, so it reshapes the spectrum of whatever drives it. A guitar body's dozens of resonances are collectively a fixed filter through which every note passes, and two guitars differ because their filters differ.

That view, a source with a spectrum, passed through a filter with a shape, is the single most useful model in the rest of this book. It is the model of the singing voice (Chapter 13), of the vowel, of subtractive synthesis (Chapter 15), and of why a violin sounds like a violin whatever note it plays.

4.6 Summary

- **Free vibration** happens at an object's natural frequency; **driven vibration** happens at the driving frequency. What the natural frequency controls is not the frequency of the response but its size.
- **Resonance** is the large response obtained by driving at a natural frequency. It works because at resonance the drive leads the displacement by exactly 90°, in phase with the velocity, so every push adds energy.
- **The quality factor** $Q = f_0/\Delta f = \pi f_0 \tau$ measures the sharpness of a resonance. A narrow peak and a long ring are the same fact stated two ways.
- **Loudness and sustain trade against each other**, because radiating sound is a loss of energy and losses lower Q. Instruments therefore want a high-Q vibrator and a low-Q resonator: a tuning fork has Q in the thousands and is nearly silent, a guitar body has Q near five and is loud.
- **A normal mode** is a pattern in which every part oscillates at one frequency with fixed relative amplitudes. A system with N degrees of freedom has exactly N of them; a continuous object has unlimited numbers.
- **Any motion is a sum of normal modes.** A plucked string is decomposed into modes, each oscillates at its own frequency, and the sum is the motion. Plucking at L/n excites no nth mode, because the pluck point is a node of it.
- **A string's modes are harmonic**, exact whole-number ratios, but this is special. Bars and membranes are not, so percussion instruments mostly lack definite pitch.
- **A Helmholtz resonator** is a plug of air on a spring of air, with a single resonance at $f_0 = (v/2\pi)\sqrt{A/VL_{\text{eff}}}$. It is a bottle, a guitar's air mode, and a loudspeaker's bass port.
- **A resonator is both an amplifier and a filter**: it speeds energy out into the air, and it reshapes the spectrum passing through it. Source-plus-filter is the model used for the rest of the book.

4.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity. The set mixes multiple response, true/false, drag-the-words, fill-in-the-blank, and mark-the-words formats.

1. **Multiple response.** Which changes lower the resonance frequency of a Helmholtz resonator? Select from: increase cavity volume; lengthen the effective neck; increase neck area; increase the speed of sound.

2. **True or false.** After transients die away, a driven oscillator vibrates at its natural frequency rather than at the driving frequency.

3. **Drag the words.** Complete: quality factor is resonance frequency divided by _____; a high-Q resonance has a _____ response peak and a _____ decay time. (Terms: bandwidth, narrow, long, wide, short.)

4. **Fill in the blank.** A system with two degrees of freedom has _____ normal modes.

5. **Mark the words.** Identify the two functions of a resonator in this sentence: "A resonator transfers energy efficiently to the air and filters the source spectrum; it does not create energy."

4.8 Conceptual Questions

1. A driven oscillator settles into oscillating at the driving frequency rather than its own. Explain what role its natural frequency then plays.

2. A child on a swing is pushed once per swing, and the swing grows. Explain why pushing twice per swing does not work, in terms of the phase of the drive relative to the velocity.

3. A tuning fork has a Q of several thousand and a guitar body a Q of about five. Which is louder, which rings longer, and explain why those two answers must go together.

4. Explain why a tall resonance peak is necessarily a narrow one, without using any formula.

5. Two masses connected by springs have two normal modes. Explain why the in-phase mode has the lower frequency.

6. A guitarist plucks a string exactly one-third of the way along. Which harmonics are missing from the resulting note, and why?

7. Explain why a bottle has essentially one resonance while an organ pipe has a whole harmonic series, even though both are containers of air.

8. In a violin, the string sets the pitch and the body responds. In a clarinet, the air column tells the reed what pitch to sound. Explain why this difference makes the two instruments require different kinds of analysis.

4.9 Problems

Exercise 4.1
(Straightforward) A resonance is measured with a peak at 440 Hz and half-power points at 426 Hz and 454 Hz. (a) What is the bandwidth? (b) What is Q? (c) How long would this resonance ring after being struck?

Solution to Exercise 4.1
(a) $\Delta f = 454 - 426 = 28$ Hz.
(b) $Q = f_0/\Delta f = 440/28 = 15.7$.
(c) From $Q = \pi f_0 \tau$:

$$\tau = \frac{Q}{\pi f_0} = \frac{15.7}{\pi(440\ \text{Hz})} = 0.0114\ \text{s}. \tag{4.13}$$

Therefore, $Q \approx 16$ and the free decay time is about 11 ms.

Exercise 4.2
(Straightforward) A wine glass rings at 720 Hz and takes 2.8 s for its amplitude to fall to $1/e$ of its initial value. (a) Find its Q. (b) Find the bandwidth of its resonance. (c) Comment on how precisely a singer would have to match the pitch to excite it.

Solution to Exercise 4.2

(a) From $Q = \pi f_0 \tau$:

$$Q = \pi(720\ \text{Hz})(2.8\ \text{s}) = 6.3 \times 10^3. \tag{4.14}$$

(b) From $Q = f_0/\Delta f$:

$$\Delta f = \frac{f_0}{Q} = \frac{720\ \text{Hz}}{6330} = 0.11\ \text{Hz}. \tag{4.15}$$

(c) The singer must land within roughly a tenth of a hertz of 720 Hz, about a quarter of a cent, and hold it. That is far finer than any singer can aim deliberately, so the demonstration is usually done by sweeping slowly through the region until the glass responds, rather than by hitting the note directly.

Exercise 4.3

(Moderate) A guitar's air resonance is at 98 Hz with $Q = 6.0$. (a) Find the bandwidth. (b) Over what range of notes does the resonance provide at least half its peak power? (c) Express that range in semitones.

Solution to Exercise 4.3

(a) $\Delta f = f_0/Q = (98\ \text{Hz})/6.0 = 16.3\ \text{Hz}$.

(b) Centered on 98 Hz, the half-power range runs from about $98 - 8.2 = 89.8$ Hz to $98 + 8.2 = 106.2$ Hz.

(c) In semitones:

$$n = 12\log_2\!\left(\frac{106.2}{89.8}\right) = 12(0.2420) = 2.9\ \text{semitones}. \tag{4.16}$$

Therefore, the resonance usefully covers about three semitones. A low Q is what makes even that possible; the wine glass of the previous problem covers about four *thousandths* of a semitone.

Exercise 4.4

(Straightforward) A bottle of internal volume 0.75 L has a neck of cross-sectional area 3.8 cm^2 and effective length 7.5 cm. Take $v = 343$ m/s. (a) What note does it sound when blown across? (b) It is half filled with water. What note does it sound now?

Solution to Exercise 4.4

(a) Converting to SI: $V = 7.5 \times 10^{-4}\ \text{m}^3$, $A = 3.8 \times 10^{-4}\ \text{m}^2$, $L_{\text{eff}} = 0.075\ \text{m}$.

$$f_0 = \frac{343\ \text{m/s}}{2\pi}\sqrt{\frac{3.8 \times 10^{-4}\ \text{m}^2}{(7.5 \times 10^{-4}\ \text{m}^3)(0.075\ \text{m})}} = (54.6\ \text{m/s})\sqrt{6.76\ \text{m}^{-2}} = (54.6\ \text{m/s})(2.60\ \text{m}^{-1}) = 142\ \text{Hz}. \tag{4.17}$$

(b) Halving the air volume, with the neck unchanged:

$$f_0' = f_0\sqrt{2} = (142\ \text{Hz})\sqrt{2} = 201\ \text{Hz}. \tag{4.18}$$

Therefore, the empty bottle sounds about 142 Hz and the half-full one about 201 Hz: a rise of $1200\log_2\sqrt{2} = 600$ cents, exactly a tritone. Halving the volume always raises the pitch by exactly this interval, whatever the bottle.

Exercise 4.5

(Straightforward) A string is plucked at exactly one-quarter of its length. (a) Which harmonics are absent? (b) If the string sounds 220 Hz, what are the frequencies of the three lowest absent harmonics? (c) Where should the player pluck to suppress the third harmonic instead?

Solution to Exercise 4.5

(a) Plucking at $L/4$ puts the pluck point at a node of the 4th mode and of every multiple of 4. So harmonics 4, 8, 12, …are absent.

(b) $4 \times 220 = 880$ Hz, $8 \times 220 = 1760$ Hz, $12 \times 220 = 2640$ Hz.

(c) At $L/3$, which is a node of the 3rd mode (and of the 6th, 9th, …).

Therefore, the pluck position acts as a comb filter on the harmonic series, removing one harmonic and all its multiples, which is exactly the control a guitarist exercises by moving the right hand.

Exercise 4.6

(Challenging) A resonator has $f_0 = 250$ Hz. It is driven at (a) 250 Hz, (b) 125 Hz, (c) 500 Hz, all with the same force amplitude. Using the response formula with $Q = 10$, find the amplitude in each case relative to the response to a very slow push.

Solution to Exercise 4.6

The response is $A/A_{\text{static}} = 1/\sqrt{(1-r^2)^2 + (r/Q)^2}$, with $r = f/f_0$.

(a) $r = 1$: the first term vanishes, leaving

$$A = \frac{1}{\sqrt{(1/10)^2}} = 10. \tag{4.19}$$

(b) $r = 0.5$:

$$A = \frac{1}{\sqrt{(1-0.25)^2 + (0.05)^2}} = \frac{1}{\sqrt{0.5625 + 0.0025}} = 1.33. \tag{4.20}$$

(c) $r = 2$:

$$A = \frac{1}{\sqrt{(1-4)^2 + (0.2)^2}} = \frac{1}{\sqrt{9 + 0.04}} = 0.333. \tag{4.21}$$

Therefore, the amplitude at resonance is $Q = 10$ times the static response, while an octave below it is only 1.3 times and an octave above it is *one third*: the resonator suppresses frequencies above its peak rather than merely failing to boost them.

Exercise 4.7

(Moderate) Two identical masses are joined by a spring of stiffness k_c and each is joined to a wall by a spring of stiffness k. The in-phase mode has frequency $f_1 = (1/2\pi)\sqrt{k/m}$ and the out-of-phase mode $f_2 = (1/2\pi)\sqrt{(k + 2k_c)/m}$. (a) If $k_c = k$, what is the ratio f_2/f_1? (b) What happens to the ratio as the coupling spring becomes very weak?

Solution to Exercise 4.7

(a) With $k_c = k$:

$$\frac{f_2}{f_1} = \sqrt{\frac{k + 2k}{k}} = \sqrt{3} = 1.73. \tag{4.22}$$

(b) As $k_c \to 0$,

$$\frac{f_2}{f_1} = \sqrt{\frac{k + 2k_c}{k}} \to 1. \tag{4.23}$$

Therefore, strong coupling splits the two mode frequencies widely apart, and weak coupling brings them together. In the limit of no coupling the two masses are independent oscillators of identical frequency: the "two modes" have become indistinguishable, which is what you would expect of two unconnected copies of the same thing.

Exercise 4.8

(Straightforward) A resonator stores 2.4 mJ of energy and loses 0.15 mJ per cycle. (a) Find its Q. (b) If its frequency is 180 Hz, how long does it ring? (c) How many cycles is that?

Solution to Exercise 4.8

(a) From $Q = 2\pi \times$ (energy stored)/(energy lost per cycle):

$$Q = 2\pi \, \frac{2.4}{0.15} = 2\pi(16) = 101. \tag{4.24}$$

(b) From $Q = \pi f_0 \tau$:

$$\tau = \frac{101}{\pi(180 \text{ Hz})} = 0.178 \text{ s}. \tag{4.25}$$

(c) The number of cycles in τ is $f_0\tau = (180 \text{ Hz})(0.178 \text{ s}) = 32$ cycles.

Therefore, $Q \approx 100$, the amplitude falls to $1/e$ in about 0.18 s, and that is about 32 cycles. Note the general rule visible here: a resonator rings for roughly Q/π cycles, whatever its frequency.

Exercise 4.9

(Moderate) An engineer wants a loudspeaker enclosure with a bass-reflex port tuned to 45 Hz. The cabinet's internal volume is 32 L and the port is a tube of radius 2.5 cm. Ignoring end corrections, what length of port is needed?

Solution to Exercise 4.9

Rearranging the Helmholtz formula for length:

$$f_0 = \frac{v}{2\pi}\sqrt{\frac{A}{VL}} \;\Rightarrow\; L = \frac{A}{V}\left(\frac{v}{2\pi f_0}\right)^2. \tag{4.26}$$

With $A = \pi(0.025)^2 = 1.963 \times 10^{-3}$ m^2, $V = 0.032$ m^3:

$$L = \frac{1.963 \times 10^{-3} \text{ m}^2}{0.032 \text{ m}^3}\left(\frac{343 \text{ m/s}}{2\pi(45 \text{ Hz})}\right)^2 = (0.0614 \text{ m}^{-1})(1.213 \text{ m})^2 = 0.0903 \text{ m}. \tag{4.27}$$

Therefore, a port about 9.0 cm long is needed. In practice the end correction would shorten this by several centimeters, which for a port this wide is a large fraction of the answer: the same warning as §3.5.

Exercise 4.10
(Challenging) Show that a resonator rings for approximately Q/π cycles, and use the result to estimate how many cycles a tuning fork of $Q = 4000$ rings for.

Solution to Exercise 4.10
The number of cycles in one decay time τ is $N = f_0\tau$. From $Q = \pi f_0 \tau$,

$$\tau = \frac{Q}{\pi f_0} \;\Rightarrow\; N = f_0 \frac{Q}{\pi f_0} = \frac{Q}{\pi}. \tag{4.28}$$

The frequency has canceled: **the number of cycles depends only on Q.**
For the tuning fork,

$$N = \frac{4000}{\pi} = 1.3 \times 10^3 \text{ cycles}. \tag{4.29}$$

Therefore, a fork of $Q = 4000$ rings for about 1300 cycles before falling to $1/e$. At 440 Hz that is about three seconds; at 256 Hz, about five. The lower fork rings *longer in time* while ringing the same number of cycles.

Exercise 4.11
(Moderate) A drumhead has modes at 1.00, 1.59, 2.14, 2.30, and 2.65 times its lowest frequency. A string has modes at 1, 2, 3, 4, 5 times its lowest. (a) Express each set as intervals in cents above the lowest mode. (b) Explain, from the two lists, why the string has a definite pitch and the drumhead does not.

Solution to Exercise 4.11
(a) Using $n = 1200 \log_2(\text{ratio})$:

Mode	Drumhead ratio	cents	String ratio	cents
1	1.00	0	1	0
2	1.59	804	2	1200
3	2.14	1318	3	1902
4	2.30	1444	4	2400
5	2.65	1688	5	2786

(b) The string's partials are exact whole-number multiples of one frequency, so they are all harmonics of a single fundamental, and the auditory system can assign one pitch to the whole set (Chapter 8). The drumhead's are not multiples of anything: 1.59, 2.14, and 2.30 have no common fundamental, so there is no single frequency the set points to.

Therefore, the drumhead gives the ear a collection of unrelated frequencies rather than a harmonic series, and the ear reports a sound with a rough pitch region rather than a note.

Exercise 4.12
(Moderate) A violin body has its main resonances near 280 Hz and 460 Hz, each with $Q \approx 25$, and the instrument's range runs from 196 Hz to about 3000 Hz. (a) Find the bandwidth of each resonance. (b) What fraction of the instrument's range is covered by the two together? (c) What does your answer imply about how the rest of the range is radiated?

Solution to Exercise 4.12

(a) $\Delta f_1 = 280/25 = 11.2$ Hz and $\Delta f_2 = 460/25 = 18.4$ Hz.

(b) Together they cover about 30 Hz out of a range of $3000 - 196 = 2804$ Hz, or about 1%.

(c) Two resonances plainly cannot account for the instrument's ability to radiate every note. The real body has *dozens* of modes above these, increasingly closely spaced, which merge into a broad and lumpy response rather than a set of isolated peaks. The two named resonances matter because they are isolated and therefore individually audible, they are the ones a maker can tune, but the rest of the range is radiated by a forest of overlapping higher modes.

Exercise 4.13

(Straightforward) A string is plucked at its exact midpoint. (a) Which harmonics are absent? (b) Compare the brightness of this note with one plucked near the bridge, and explain the difference in terms of the mode amplitudes.

Solution to Exercise 4.13

(a) The midpoint is a node of every even mode, so harmonics 2, 4, 6, 8, …are absent. Only odd harmonics remain.

(b) A midpoint pluck is the *least* bright pluck available. It removes every even harmonic outright, and the $\sin(n\pi a/L)$ factor together with the $1/n^2$ falloff leaves the surviving odd harmonics weak as well. A pluck close to the bridge, at small a, makes $\sin(n\pi a/L)$ small for low n and comparatively large for high n, so the spectrum tilts toward the high harmonics and the note is bright and thin.

Therefore, the same string can be made mellow or brilliant by nothing more than the choice of pluck point: a guitarist's right hand moves as much as the left for that reason.

Exercise 4.14

(Moderate) A maker is deciding how heavily to brace a guitar top. Heavier bracing raises the body's Q from 5 to 9. (a) By what factor does the peak response change, at fixed damping-independent drive? (b) By what factor does the bandwidth change? (c) State in words what the player would notice.

Solution to Exercise 4.14

(a) The peak response at resonance is proportional to Q, so it rises by $9/5 = 1.8$.

(b) The bandwidth is f_0/Q, so it falls by the same factor: to $5/9 = 0.56$ of what it was.

(c) The player would find that notes near the body resonance are louder and ring longer, while notes away from it are relatively weaker: the instrument has become less even across its range, and more colored. It would also sustain more and project less overall, since a higher Q means less energy is leaving as sound per cycle.

Therefore, the trade-off of §4.3 appears directly as a design decision: bracing for evenness and volume means accepting a shorter sustain, and bracing for sustain means accepting an uneven instrument.

Part II

Spectrum and Timbre

Chapter 5

Fourier Analysis, Harmonics, and Timbre

Learning Objectives

By the end of this chapter, you should be able to:

- State Fourier's theorem for a periodic signal, and explain what it guarantees about any sustained musical tone.
- Build a sawtooth, square, and triangle wave by adding harmonics, and predict the effect on the waveform of adding or removing a given partial.
- Read a line spectrum: identify the fundamental, name the partials by harmonic number, and relate the spectrum to the waveform beside it.
- Distinguish harmonic from partial from overtone, and use each term correctly.
- Explain why two tones of the same pitch and loudness can sound entirely different, and identify the spectral features responsible.
- Explain why the phase relationships among harmonics change the waveform dramatically but the timbre hardly at all, and state the limits of that rule.
- Describe the attack, decay, sustain, and release of a note, and give evidence that transients matter at least as much as the steady-state spectrum for recognizing an instrument.
- Read a spectrogram, and use it to describe how the spectrum of a real note changes over its duration.
- Distinguish harmonic spectra from inharmonic ones and from noise, and identify which instruments produce each.

Introduction

A clarinet and a violin play the same note, at the same loudness, for the same length of time. Nobody has any difficulty telling them apart.

That is a problem for everything the book has said so far. Chapter 1 established that what reaches your ear is a single wiggling number, pressure at a point, as a function of time. If the pitch is the same and the loudness is the same, what is left for the difference to live in?

The answer is the **shape** of the wiggle, and the tool that makes shape analyzable is the subject of this chapter.

Jean-Baptiste Joseph Fourier, working on heat conduction in the 1800s, proved something that at first seems too strong to be true: **any periodic waveform, however jagged, is a**

sum of sine waves at whole-number multiples of one frequency. Not approximately, exactly, and in only one way.

Applied to music, this says that a sustained note is completely specified by a list: how much of each harmonic it contains. That list is the **spectrum**, and it is the single most useful object in the rest of this book. Chapters 10 through 13 are, in the end, four long answers to the question "what spectrum does this instrument produce, and why?"

This chapter also delivers the first serious complication. The spectrum is not the whole story, and the part it leaves out, how the sound *starts*, turns out to matter at least as much.

5.1 Fourier's Theorem

Any Periodic Wave Is a Sum of Sinusoids

The statement, precisely: if a function repeats with period T, it can be written as

$$y(t) = A_1 \sin(2\pi f_1 t + \phi_1) + A_2 \sin(2\pi \cdot 2 f_1 t + \phi_2) + A_3 \sin(2\pi \cdot 3 f_1 t + \phi_3) + \cdots \tag{5.1}$$

where $f_1 = 1/T$. Every component is a **harmonic** of f_1: its frequency is an exact whole-number multiple.

Two features of this deserve emphasis.

Only harmonics appear. Not arbitrary frequencies, multiples of one fundamental. This is forced by the repetition: any component at a non-multiple frequency would not come back to where it started after time T, and so the whole would not repeat.

The decomposition is unique. One waveform, one list of amplitudes and phases. There is no ambiguity about what a sound "is made of".

Building a Sawtooth by Hand

The theorem is easier to believe once you have watched it happen.

That figure should be uncomfortable the first time. The building blocks are smooth, endless, curvy sinusoids; the target has sharp corners and instantaneous jumps. Adding smooth things and getting a corner seems like cheating. It is not: the corner is what infinitely many harmonics, in exactly the right proportions, add up to.

The Fourier coefficients, as integrals

Finding the amplitudes is the same trick as §4.4 used for a plucked string, and for the same reason: sinusoids of different harmonic numbers are **orthogonal** over one period.

Write the series in the equivalent form

$$y(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \Big[a_n \cos(2\pi n f_1 t) + b_n \sin(2\pi n f_1 t) \Big]. \tag{5.2}$$

The orthogonality relations are that, over one period T,

$$\int_0^T \cos(2\pi n f_1 t) \cos(2\pi m f_1 t)\, \mathrm{d}t = \begin{cases} T/2 & n = m \neq 0 \\ 0 & n \neq m, \end{cases} \tag{5.3}$$

with the same for sines, and $\int_0^T \cos(2\pi n f_1 t) \sin(2\pi m f_1 t)\, \mathrm{d}t = 0$ for all n and m.

So multiply the series by $\cos(2\pi m f_1 t)$ and integrate. Every term dies except one, leaving

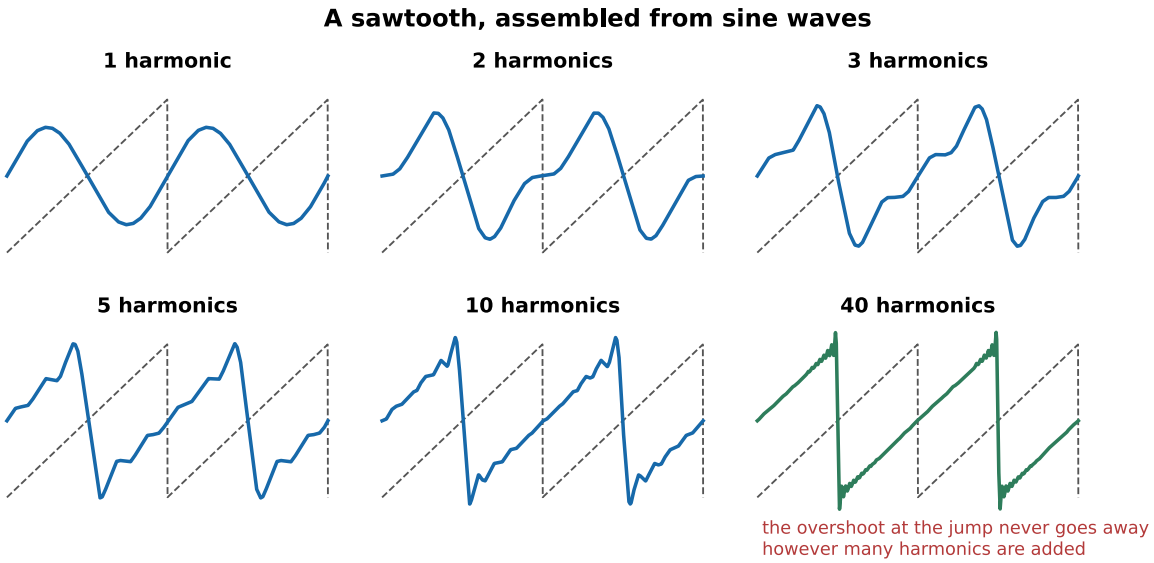


Figure 5.1: A sawtooth wave, assembled from sine waves whose amplitudes fall as $1/n$. One harmonic is a sine; by ten it is recognizably a sawtooth; by forty the approximation is very close. The overshoot at each vertical jump, however, never disappears, it only gets narrower. That is the **Gibbs phenomenon**, and it is a genuine property of the sum rather than an error in the drawing.

$$a_m = \frac{2}{T}\int_0^T y(t)\cos(2\pi m f_1 t)\,\mathrm{d}t, \qquad b_m = \frac{2}{T}\int_0^T y(t)\sin(2\pi m f_1 t)\,\mathrm{d}t. \tag{5.4}$$

The amplitude and phase of the mth harmonic are then $A_m = \sqrt{a_m^2 + b_m^2}$ and $\phi_m = \arctan(a_m/b_m)$.

For the sawtooth $y(t) = 2(t/T) - 1$ on $0 \le t < T$, the integrals give $a_m = 0$ and $b_m = -2/(\pi m)$: every harmonic present, falling as $1/n$, which is what the figure above plots.

The general rule worth carrying away is that **the sharper the waveform, the more slowly its harmonics die off**. A discontinuity gives $1/n$; a kink in the slope gives $1/n^2$; a perfectly smooth wave has harmonics that vanish faster than any power. This is why sharp-cornered waveforms sound bright.

Analysis and Synthesis

The theorem runs in both directions, and both are used constantly.

Analysis takes a sound and finds its spectrum. This is what a spectrum analyzer does, what the fast Fourier transform computes, and what the laboratory exercises in Laboratory Exercises ask you to do to real instruments.

Synthesis takes a spectrum and builds the sound. This is additive synthesis (Chapter 15), and it is how every audio example in this book was made.

The two are exact inverses, which is a stronger claim than it may look. It means the spectrum contains *everything* about a periodic sound: no information is lost in going from the waveform to the list of harmonics and back again.

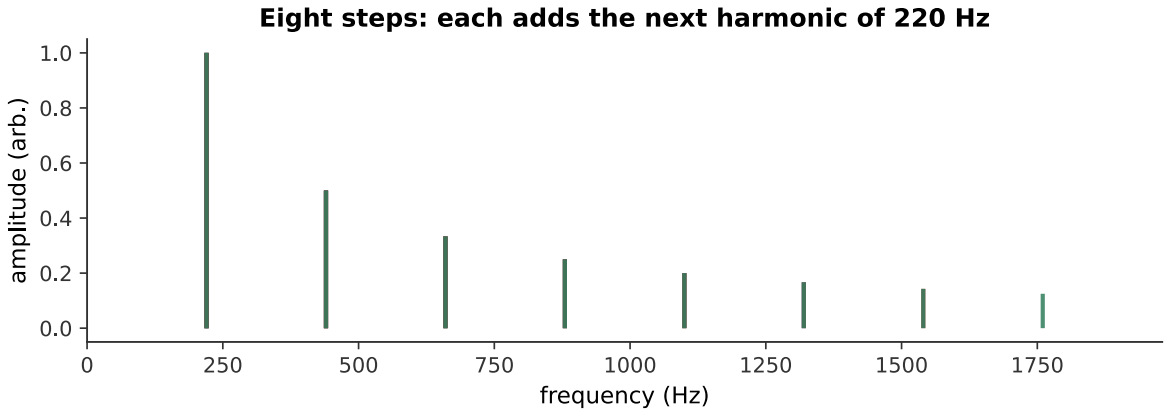


Figure 5.2: Synthesis, one harmonic at a time. Each step adds the next harmonic of 220 Hz at the amplitude a sawtooth calls for. Listen for what *does not* happen: the pitch never moves. Adding higher frequencies to a tone makes it brighter, not higher, which is the first and most important thing the ear does with a harmonic series. *What you hear:*A tone that brightens in eight steps, growing from a plain hollow sound to a buzzy one. The pitch does not change at any point. Listen: Harmonic Build Up

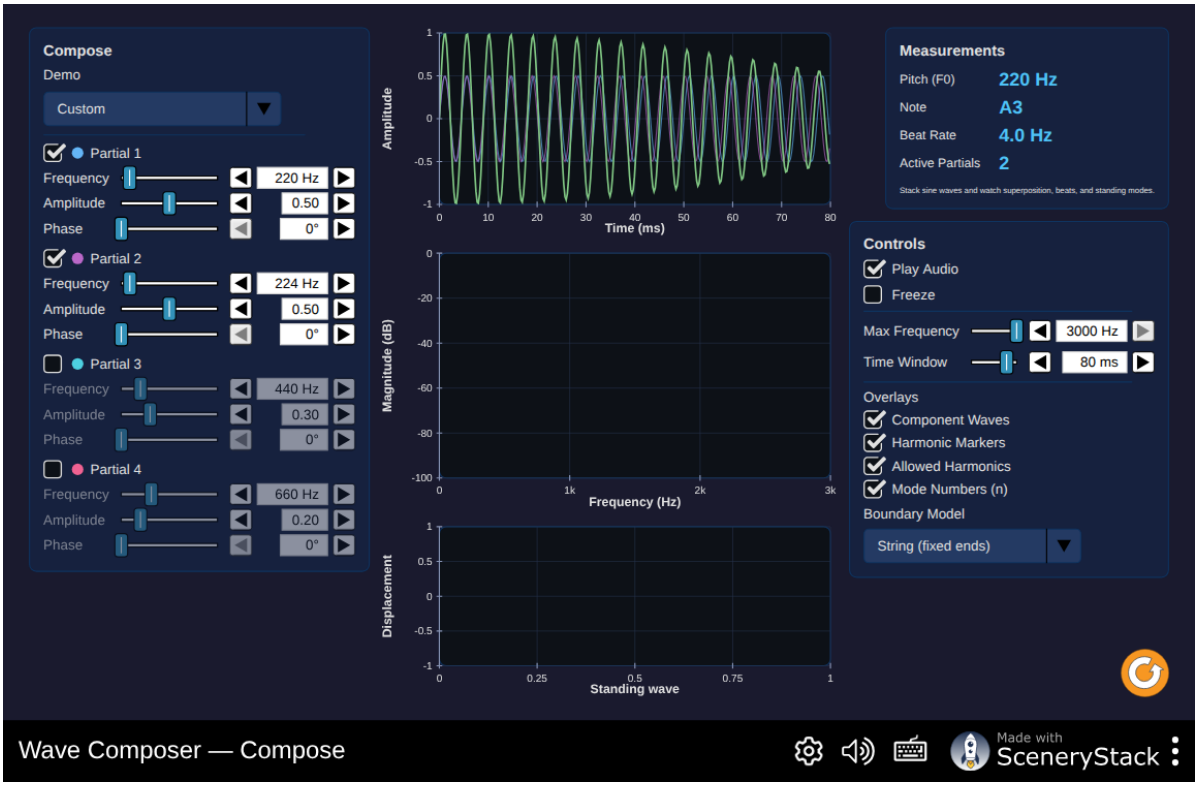


Figure 5.3: The decomposer screen: hand it an arbitrary periodic wave and watch it resolve into a stack of harmonic amplitudes in real time. This is analysis running the other direction from the sawtooth built by hand above, spectrum from waveform rather than waveform from spectrum, and the two screens of this simulation are the theorem's two directions made concrete. Interactive simulation: Wave Composer

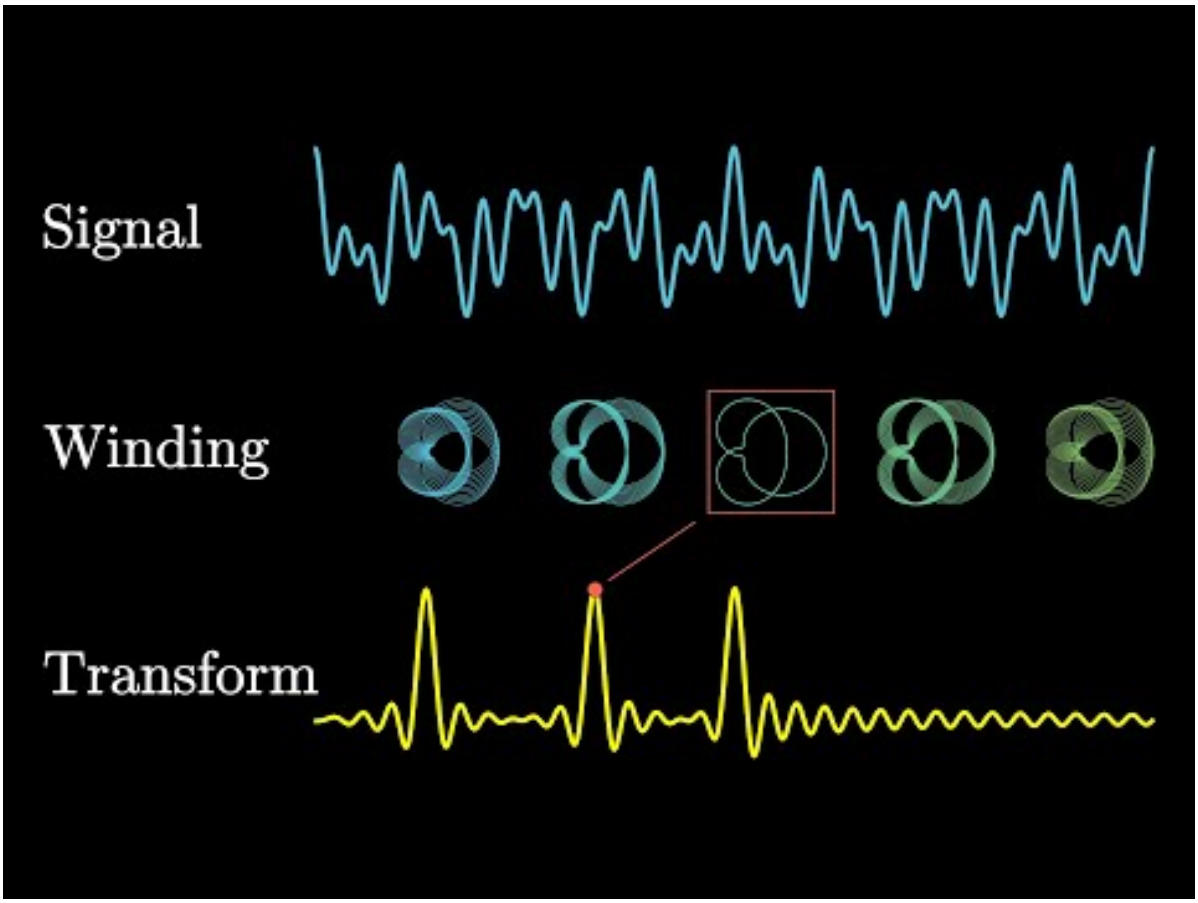


Figure 5.4: Grant Sanderson builds the Fourier transform visually by winding a waveform around a circle. The construction shows how a spectrum measures the amount of each frequency in a signal before the chapter turns to line spectra and timbre. Watch on YouTube: But What Is the Fourier Transform? A Visual Introduction

5.2 The Harmonic Series and the Spectrum

Harmonics, Partials, and Overtones

Three words, frequently confused, and worth fixing now.

- A **partial** is any frequency component of a complex tone. It is the general word and it carries no assumption.
- A **harmonic** is a partial whose frequency is an exact whole-number multiple of the fundamental. The nth harmonic has frequency nf_1. The fundamental is the 1st harmonic.
- An **overtone** is any partial above the fundamental. The 1st overtone is the 2nd harmonic.

That last off-by-one is a standing trap, and the reason this book avoids the word "overtone" except when quoting someone who used it.

The distinction between *partial* and *harmonic* is not pedantry, because plenty of musical sounds have partials that are not harmonics. A bell, a drum, a cymbal, and, by a small but audible amount, a piano string all produce partials at frequencies that are not whole-number multiples of anything. Section 5.5 and Chapter 12 take those up.

Reading a Line Spectrum

A spectrum is conventionally drawn as a **line spectrum**: a vertical stem at each harmonic, whose height is that harmonic's amplitude. Stems, not a filled curve, because in a periodic sound there is genuinely *nothing* between the harmonics, and a filled curve would suggest otherwise.

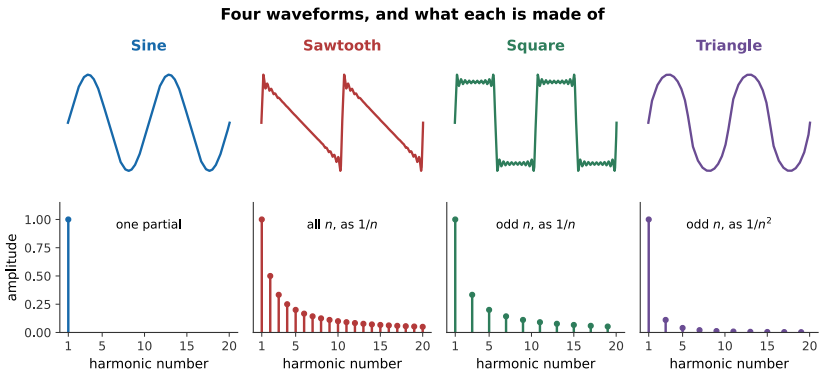


Figure 5.5: Four waveforms and their spectra. The **sawtooth** contains every harmonic, falling as $1/n$. The **square** contains only odd harmonics, also falling as $1/n$, and sounds hollow, exactly like the stopped pipe of Chapter 3, for exactly the same reason. The **triangle** contains only odd harmonics but they fall as $1/n^2$, which makes it much closer to a sine in character.

The vertical axis is often drawn in **decibels** rather than linear amplitude, for the reason Chapter 7 will give: the ear's response is logarithmic, and a partial 40 dB below the fundamental, a hundredth of its amplitude, is still perfectly audible and can matter a great deal to the timbre. On a linear plot it would be invisible.

Why Periodic Means Harmonic

Why do the partials of a sustained musical tone come out as exact multiples?

A vibrating string has modes at nf_1 because of its boundary conditions (Chapter 3). But the deeper reason applies to any sustained tone, whatever produces it: **if the waveform repeats exactly, its partials must be exact harmonics**, because anything else would prevent the repetition.

This works in reverse too, and it is the key to Chapter 8. When the ear encounters a set of partials at 200, 300, 400, 500 Hz, it can infer that the waveform repeats 100 times a second, and it reports a pitch of 100 Hz, whether or not any energy is present at 100 Hz at all.

5.3 Spectrum and Timbre

The Same Pitch, Different Sounds

Now the chapter's opening question can be answered. A clarinet and a violin playing the same note differ in their spectra, and that is what you hear.

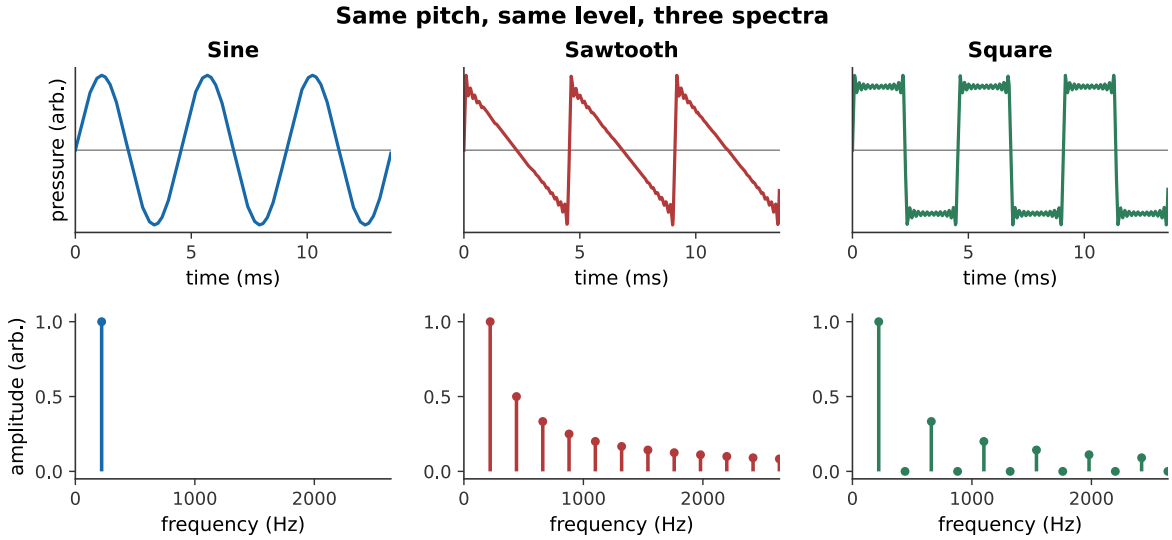


Figure 5.6: The same fundamental, the same loudness, three different sounds. Nothing separates them but which harmonics are present and how strong each one is. This is the whole of the chapter in twelve seconds. *What you hear:*Three tones of the same pitch and the same level. The sine is hollow and plain, the sawtooth buzzy and bright, the square hollow but reedy. Listen: Sine, Sawtooth, Square

Ideal waveforms make the rule clean, but instruments make it musical. The next example holds the performed pitch fixed while changing the entire mechanism that produces and filters it.

Some rough rules connect a spectrum to what it sounds like:

Spectral feature	Perceived as
Strong high harmonics	Bright, brilliant, harsh
Harmonics dying quickly	Dark, mellow, round
Odd harmonics only	Hollow, woody, reedy
Very few harmonics	Pure, thin, flute-like
Many closely spaced inharmonic partials	Clangorous, metallic

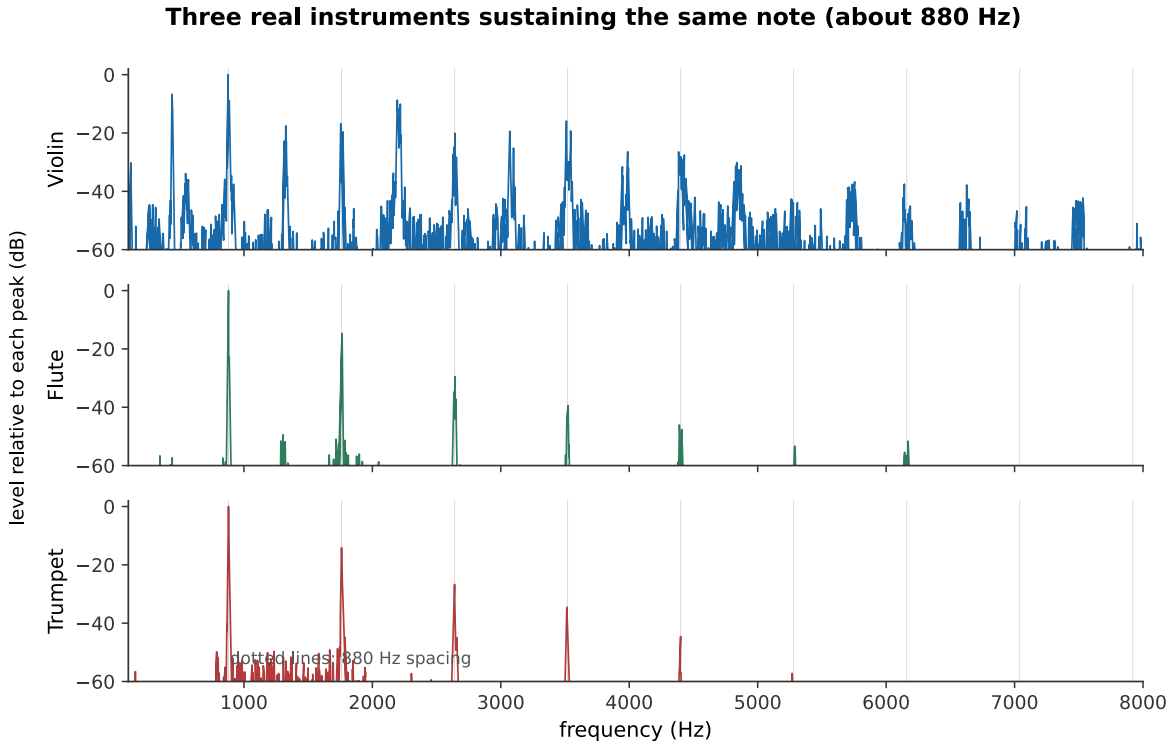


Figure 5.7: Three real instruments, one nominal fundamental near 880 Hz. Their spectral lines share the same spacing, but their envelopes, noise, attacks, and small performance fluctuations differ. These CC0 recordings are from the VSCO Community Edition; the exact source files are recorded in `SOURCES.md`. *What you hear:*Three musicians sustain the same high note. The violin is textured and continuously wavering, the flute is breathy and comparatively pure, and the trumpet is bright and brassy with strong upper harmonics. Listen: Violin, Flute, Trumpet

The **spectral centroid**, the amplitude-weighted average frequency of the partials, correlates well with judgments of brightness, and is the single number most often used when a machine has to guess what a sound is like.

Spectral Centroid, Brightness, and Roll-off

Two cautions about that correlation, because it is easy to lean on it too hard.

First, the centroid is a *summary*. Two spectra with the same centroid can sound quite different if one has its energy concentrated in a narrow band and the other spreads it widely.

Second, and more interesting: **the ear is much more sensitive to the presence or absence of a harmonic than to its exact level.** Removing every even harmonic from a tone, turning a sawtooth into a square, is instantly audible and transforms the character of the sound. Halving the amplitude of the seventh harmonic alone is barely audible at all. The pattern matters more than the precise numbers.

The Phase Question

Fourier's theorem gives each harmonic an amplitude *and* a phase. What does the phase do?

To the *waveform*, a great deal. Change the phases and the shape of the wave changes completely.

To the *ear*, remarkably little.

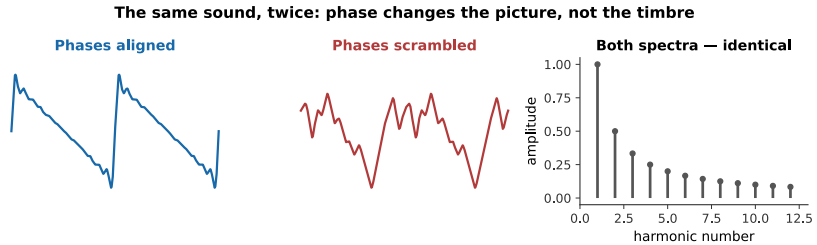


Figure 5.8: Two signals with identical spectra and randomized relative phases. The waveforms look nothing like each other. They sound almost exactly the same.

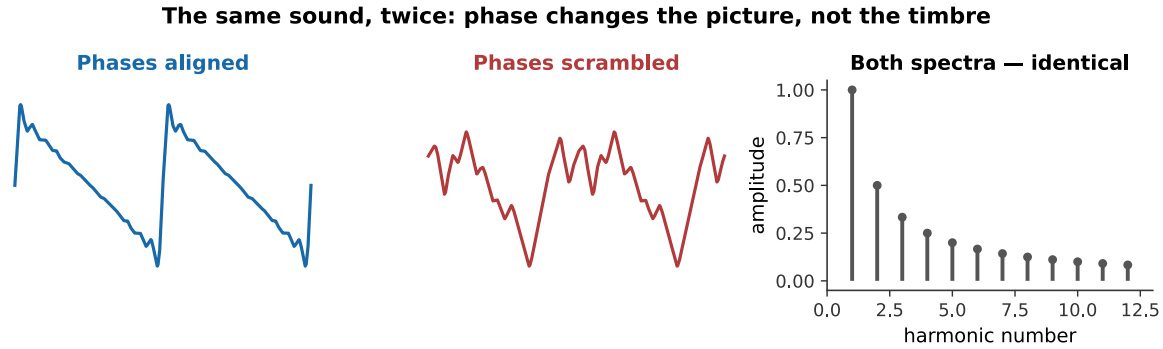


Figure 5.9: The two signals of the figure above. Try to hear a difference before reading on. This is **Ohm's acoustic law**: the ear analyzes a steady sound into its component frequencies and is largely deaf to the phase relationships between them. *What you hear:*Two buzzy tones. They are very hard to tell apart, at most there is a slight difference in the quality of the buzz, and many listeners hear none at all. Listen: Phases aligned, Phases scrambled

This is a genuinely strange fact, and it has two important consequences.

It means the ear is doing something like a Fourier analysis and then discarding half of the result. Chapter 6 shows the mechanism: the cochlea sorts frequencies by *place*, and place is preserved while relative timing between widely separated frequencies is not.

It also means audio engineers can be relaxed about phase in ways they cannot be about amplitude. A loudspeaker that shifts the phases of different frequencies by different amounts, and all of them do, is not thereby ruining the sound.

Warning

The rule has limits, and they are worth knowing. Phase matters when it changes the *envelope*: two partials very close in frequency beat against each other, and their relative phase determines when. Phase matters between the two ears, where it is a primary cue for direction (Chapter 15). And extreme phase manipulation is audible on transient-rich material. What Ohm's law says is that for a **steady** tone, phase is nearly inaudible, not that phase never matters.

5.4 Beyond the Steady State

Attack, Decay, Sustain, Release

Everything so far has assumed a sound that goes on forever. Real notes start and stop, and the way they do it is described by the **envelope**, conventionally divided into four stages.

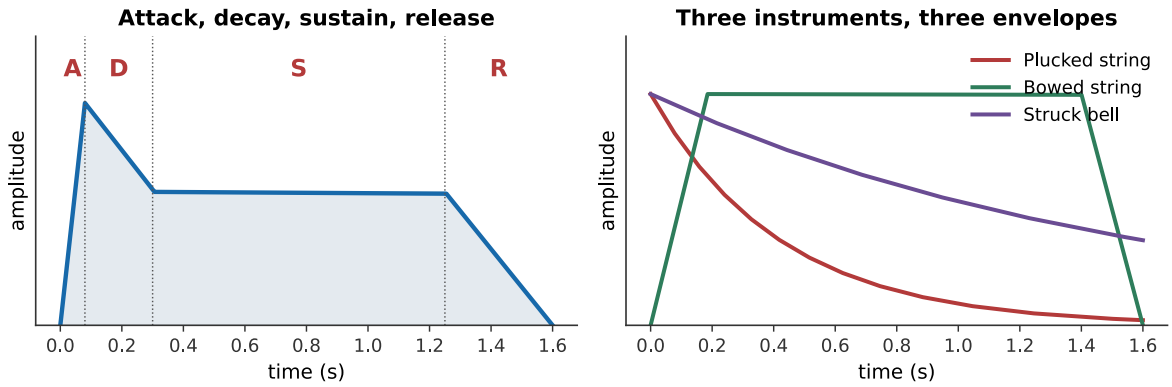


Figure 5.10: Left: the idealized four-stage envelope, borrowed from synthesizer design. Right: three real shapes. A plucked string has essentially no sustain, it starts decaying immediately. A bowed string has no decay at all while the bow keeps moving. A struck bell is a plucked string with a much longer decay.

The envelope alone distinguishes whole families of instrument. Anything struck or plucked must decay, because no energy is being added after the initial event. Anything bowed, blown, or sung can sustain indefinitely, because the player keeps supplying energy. That division, **impulsive** against **sustained** excitation, organizes Part V as much as the choice of vibrator does.

Transients Carry the Identity

Here is the result that unsettles the neat picture of §5.3.

Record a note on a real instrument. Cut off the first tenth of a second. Play the remainder to a listener and ask them to name the instrument.

They will struggle. Piano, guitar, and bowed notes with their attacks removed are surprisingly easy to confuse, and a piano with its attack removed does not sound much like a piano at all. The steady-state spectrum, the thing this chapter has spent five sections on, turns out to be a weaker cue than the first fraction of a second.

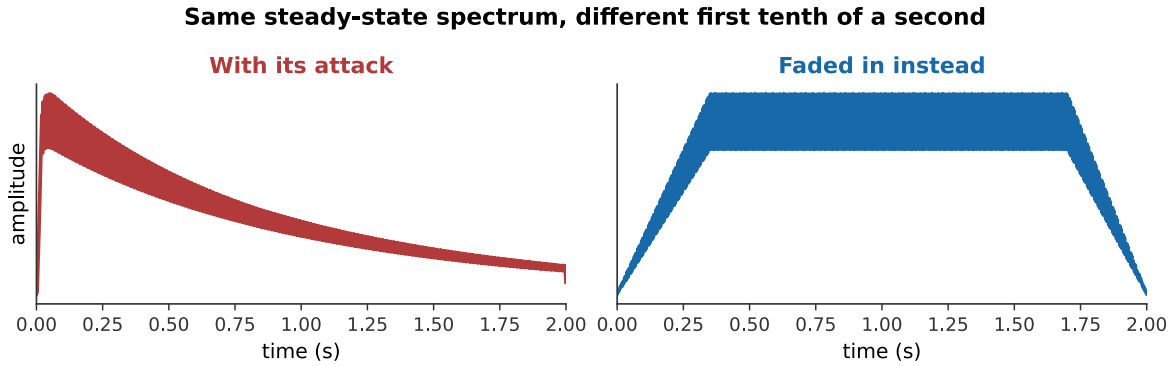


Figure 5.11: The same steady-state spectrum, with and without a struck attack. Nothing about the harmonic content differs. What differs is the first tenth of a second, and it is enough to move the sound into a different instrumental family. *What you hear:*Two notes with the same pitch and the same harmonics. The first begins with a sharp, percussive onset and sounds struck; the second fades in smoothly and sounds like an organ or a synthesizer, despite being made of exactly the same partials. Listen: With its attack, Faded in instead

What is happening in that first tenth of a second? Several things at once: the partials do not all arrive together, they do not all arrive at their final amplitudes, and there is usually a burst of **noise**: the scrape of a bow catching, the breath of a flute, the thump of a hammer, the click of a key. None of it is periodic, so none of it is in the spectrum, and all of it is informative.

The Spectrogram

A single spectrum describes a sound that is not changing. Real sounds change constantly, and the tool for those is the **spectrogram**: a picture with time across the bottom, frequency up the side, and intensity shown as brightness.

The spectrogram is made by chopping the signal into short overlapping windows and taking the spectrum of each. That construction contains an unavoidable trade-off, which Chapter 15 meets again in a different guise: a **short** window locates events precisely in time but cannot resolve closely spaced frequencies, and a **long** window does the opposite. You cannot have both, and the choice of window length is the first thing to check when reading someone else's spectrogram.

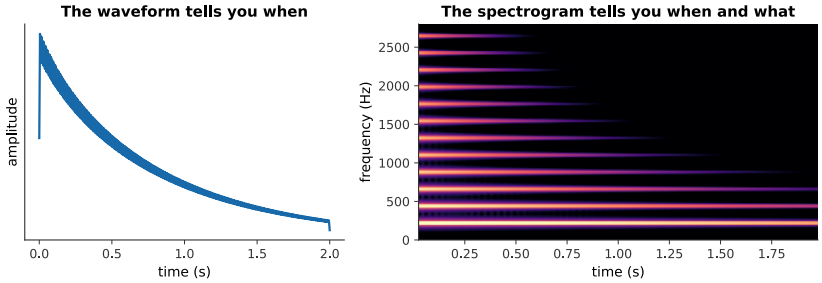


Figure 5.12: A plucked note, twice. The waveform on the left shows *when* things happen. The spectrogram on the right shows when **and what**: each horizontal band is a harmonic, and the higher bands fade first. This is the general behavior of struck and plucked instruments, and it is why a note gets darker as it dies away.

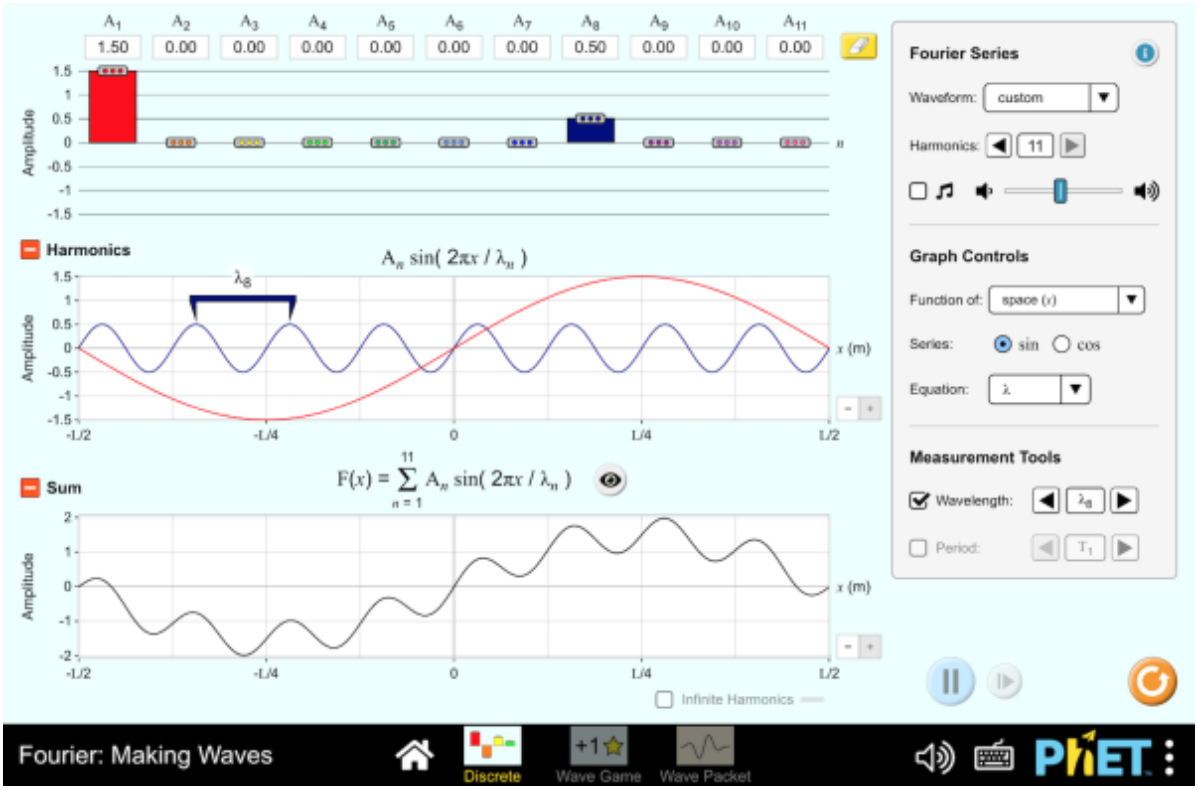


Figure 5.13: Build a waveform by dragging the amplitude of each harmonic and watch the shape respond. Two experiments repay the effort: reproduce the square wave by setting the even harmonics to zero, and then change the *phase* of a single harmonic and watch the waveform transform while the spectrum does not move at all. Interactive simulation: Fourier Making Waves

5.5 When the Spectrum Is Not Harmonic

Inharmonicity

A real piano string is not an ideal string. It has **stiffness**, it resists being bent, not merely being stretched, and stiffness adds a restoring force that grows with curvature. High modes are more curved than low ones, so they are stiffened more, and their frequencies are pushed up:

$$f_n = n f_1 \sqrt{1 + Bn^2}, \tag{5.5}$$

where B is a small **inharmonicity coefficient**, typically around 10^{-4} for a piano's middle register.

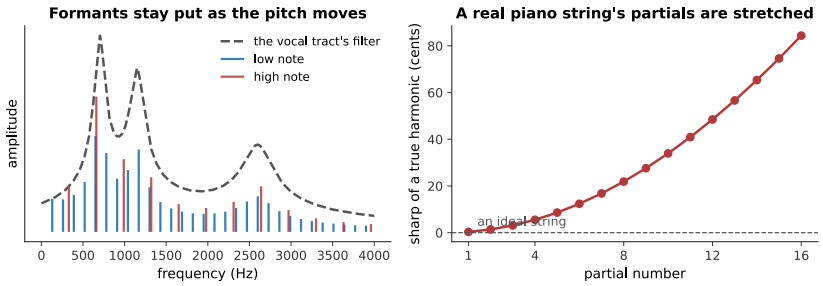


Figure 5.14: Left: **formants**, resonances fixed by the instrument's body, which shape whichever harmonics happen to fall near them. Right: **inharmonicity**, the partials of a real piano string, stretched progressively sharp of exact multiples. By the sixteenth partial the departure is most of a semitone.

The consequence is audible and important: **a piano's octaves must be stretched.** If the upper octave were tuned to exactly twice the lower fundamental, it would beat against the lower note's stretched second partial. Tuners tune the octave to the partial instead, which makes the top of a piano sharp and the bottom flat relative to equal temperament, by as much as 30 cents at the extremes. Chapter 10 works this through.

Worked example: stretching an octave

A piano's C3 string sounds $f_1 = 131$ Hz and has an inharmonicity coefficient $B = 4.0 \times 10^{-4}$, typical of a wound bass string. C4 is tuned to exactly twice this frequency, 262 Hz. Find the beat rate between C4's fundamental and C3's stretched 2nd partial, and how many cents sharp C4 would need to be tuned to eliminate it.

C3's actual 2nd partial, from $f_n = n f_1 \sqrt{1 + Bn^2}$ with $n = 2$:

$$f_2 = 2(131 \text{ Hz})\sqrt{1 + (4.0 \times 10^{-4})(4)} = (262 \text{ Hz})\sqrt{1.0016} = 262.21 \text{ Hz}. \tag{5.6}$$

Against C4's exact 262 Hz, the two beat at

$$262.21 - 262 = 0.21 \text{ Hz}, \tag{5.7}$$

about one beat every five seconds: slow enough to hear clearly, and slow enough for an impatient listener to mistake for no beat at all. Eliminating it means retuning C4 up to 262.21 Hz, which in cents is

$$1200 \log_2\left(\frac{262.21}{262}\right) = 1.4 \text{ cents.} \tag{5.8}$$

Therefore, a tuner stretches C4 by only about a cent and a half to silence this one beat, far below what anyone would notice as mistuning on its own. But every octave up the keyboard adds another such correction on top of the last, and they accumulate: by the top of the piano the total stretch reaches the tens of cents quoted above, all from the same small effect applied string after string.

Formants: Fixed Resonances, Moving Fundamental

The left panel of that figure shows the other way a spectrum acquires structure.

So far the spectrum has been described as a property of the *source*. But every instrument has a resonator (Chapter 4), and the resonator is a filter with its own fixed peaks. Those peaks are **formants**, and they emphasize whichever harmonics fall near them.

The crucial point is that **formants do not move when the pitch does**. Sing "ah" on a low note and then a high one: the fundamental changes, the harmonics change, but the resonances of your vocal tract stay where they are, because your throat has not changed shape. That is why the vowel remains recognizable, and it is the entire basis of Chapter 13.

This is the **source–filter model** of §4.5, now stated spectrally: the source supplies a harmonic series, the filter shapes it, and what you hear is the product. It applies to the voice, to every wind instrument, to a violin body, and to a subtractive synthesizer.

Noise and Its Musical Uses

Finally, the sounds Fourier's theorem does not cover.

A periodic sound has a line spectrum. A sound that never repeats has a **continuous** spectrum, energy at all frequencies rather than at a discrete set. That is **noise**, and music is full of it: the breath in a flute tone, the scrape of rosin, the consonants in sung text, the wash of a cymbal, the snares on a snare drum.

Noise is not a defect. Synthesized instruments that get the harmonic spectrum exactly right and leave the noise out sound sterile, and the single most effective improvement to an early digital instrument is usually to put the breath back. What a listener recognizes as "a real flute" is partly the harmonic series and partly the hiss that accompanies it.

There is a continuum here rather than a boundary. A cymbal's partials are so numerous and so closely spaced that the ear cannot separate them, and hears noise. A slightly unsteady sung note has partials that wander, smearing each line into a band. Chapter 12 works along this continuum from the nearly harmonic to the frankly noisy.

5.6 Summary

- **Fourier's theorem**: any periodic waveform is a sum of sinusoids at whole-number multiples of one fundamental frequency, uniquely. Periodicity forces the partials to be exact harmonics.
- **The spectrum**, the list of harmonic amplitudes, is drawn as a line spectrum, because a periodic sound genuinely has nothing between its harmonics. A decibel axis is usually more informative than a linear one.

- **Sharper waveforms have slower-decaying harmonics.** A discontinuity gives $1/n$, a kink gives $1/n^2$. This is why sharp-cornered waveforms sound bright.
- **Timbre lives in the spectrum**: same pitch, same loudness, different harmonic content, different sound. Odd harmonics alone sound hollow; strong high harmonics sound bright; few harmonics sound pure.
- **Phase changes the waveform but barely changes the sound** (Ohm's acoustic law), for a steady tone. The exceptions are beating partials, binaural cues, and transient-rich material.
- **The envelope**, attack, decay, sustain, release, divides instruments into impulsive and sustained families, and **transients carry an instrument's identity** at least as strongly as its steady-state spectrum does. Removing a note's attack makes it hard to name the instrument.
- **A spectrogram** shows spectrum against time, at the cost of an unavoidable trade-off: short windows resolve time, long windows resolve frequency, and no window does both.
- **Inharmonicity**, caused by string stiffness, stretches a piano's partials as $f_n = nf_1\sqrt{1+Bn^2}$, which forces stretched tuning.
- **Formants** are the fixed resonances of an instrument's body. They stay put as the pitch moves, so a vowel stays recognizable across a singer's range. Source-plus-filter is the model the rest of the book uses.
- **Noise** has a continuous spectrum and is a genuine part of musical sound, not a defect.

5.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple response.** Which statements describe a periodic musical waveform? Consider harmonic spacing, line spectra, phase, and odd versus even harmonics.

2. **True or false.** For a steady complex tone, changing only the harmonic phases usually changes the perceived timbre dramatically.

3. **Drag the words.** Match a waveform discontinuity and a kink to $1/n$ or $1/n^2$ spectral falloff, then match short and long analysis windows to time or frequency resolution.

4. **Fill in the blanks.** Any periodic waveform can be written as a sum of _____ at whole-number multiples of one _____ frequency.

5. **Mark the words.** Identify the four stages of an ADSR envelope: "Attack, decay, sustain, and release describe how a note changes through time; spectral centroid and formants describe its spectrum."

5.8 Conceptual Questions

1. Explain why the partials of a sustained musical tone must be exact whole-number multiples of the fundamental, without referring to any particular instrument.

2. A square wave and a sawtooth wave at the same pitch sound clearly different. State the difference in their spectra and predict which sounds hollower.

3. Explain why a waveform with a sharp corner has more high-frequency content than a smooth one.

4. Two signals have identical spectra but scrambled relative phases, and look completely different on an oscilloscope. Explain why they sound nearly the same, and name one situation in which they would not.

5. Explain why cutting the first tenth of a second off a recorded piano note makes it difficult to identify the instrument, given that the spectrum of the remainder is unchanged.

6. A spectrogram of a plucked note shows the upper bands fading before the lower ones. Explain what this says about how the note's timbre changes as it decays.

7. Explain what a formant is, and why singing a vowel at a higher pitch does not move it.

8. A cymbal has a spectrum with no clear lines in it. Explain why it has no definite pitch, and relate your answer to the modes of a plate.

5.9 Problems

Exercise 5.1
(Straightforward) A periodic waveform repeats every 4.55 ms. (a) What is its fundamental frequency? (b) What are the frequencies of its first five harmonics?

Solution to Exercise 5.1
(a) The fundamental is the reciprocal of the period:

$$f_1 = \frac{1}{4.55 \times 10^{-3}\ \text{s}} = 220\ \text{Hz}. \tag{5.9}$$

(b) Harmonics are whole-number multiples:

$$220,\ 440,\ 660,\ 880,\ 1100\ \text{Hz}. \tag{5.10}$$

Therefore, the tone is A3 and its first five harmonics run up to 1100 Hz, spanning, in musical terms, two octaves and a major third above the fundamental.

Exercise 5.2
(Moderate) A sawtooth wave has harmonic amplitudes proportional to $1/n$. (a) What is the amplitude of the 8th harmonic relative to the fundamental? (b) Express that ratio in decibels. (c) Repeat for a triangle wave, whose odd harmonics fall as $1/n^2$.

Solution to Exercise 5.2
(a) For the sawtooth, $A_8/A_1 = 1/8 = 0.125$.
(b) In decibels, using $20\log_{10}$ for an amplitude ratio:

$$20\log_{10}(0.125) = -18.1\ \text{dB}. \tag{5.11}$$

(c) For the triangle, $A_8 = 0$: the 8th harmonic is even and therefore absent. Taking the 7th instead:

$$\frac{A_7}{A_1} = \frac{1}{49} = 0.0204, \qquad 20\log_{10}(0.0204) = -33.8\ \text{dB}. \tag{5.12}$$

Therefore, the sawtooth's 8th harmonic is 18 dB down and the triangle's 7th is 34 dB down: the triangle therefore sounds much closer to a pure tone.

Exercise 5.3

(Straightforward) A tone contains partials at 300, 400, 500, 600, and 700 Hz. (a) What is the fundamental frequency? (b) Which harmonic numbers are these? (c) Is any energy present at the fundamental?

Solution to Exercise 5.3

(a) The partials are spaced 100 Hz apart and each is a multiple of 100 Hz, so $f_1 = 100$ Hz.

(b) They are the 3rd, 4th, 5th, 6th, and 7th harmonics.

(c) No, there is nothing at 100 Hz, and nothing at 200 Hz either.

Therefore, the waveform repeats 100 times a second and a listener hears a pitch of 100 Hz despite the complete absence of energy there. This is the **missing fundamental**, taken up in Chapter 8.

Exercise 5.4

(Moderate) A piano string sounds 262 Hz and has an inharmonicity coefficient $B = 1.3 \times 10^{-4}$. (a) Find the frequency of its 2nd partial. (b) Find its 8th. (c) Express each as cents sharp of a true harmonic.

Solution to Exercise 5.4

Using $f_n = n f_1 \sqrt{1 + Bn^2}$.

(a) For $n = 2$:

$$f_2 = 2(262)\sqrt{1 + (1.3 \times 10^{-4})(4)} = 524\sqrt{1.00052} = 524.14 \text{ Hz}, \tag{5.13}$$

against an exact 524 Hz. In cents:

$$1200 \log_2(524.14/524) = 0.45 \text{ cents}. \tag{5.14}$$

(b) For $n = 8$:

$$f_8 = 8(262)\sqrt{1 + (1.3 \times 10^{-4})(64)} = 2096\sqrt{1.00832} = 2104.7 \text{ Hz}, \tag{5.15}$$

against an exact 2096 Hz:

$$1200 \log_2(2104.7/2096) = 7.2 \text{ cents}. \tag{5.16}$$

Therefore, the 2nd partial is sharp by half a cent, inaudible alone, and the 8th by 7 cents, which is easily enough to produce audible beating against another string's true harmonic. That beating is what forces stretched tuning.

Exercise 5.5

(Moderate) A tone is built from harmonics of 200 Hz with amplitudes 1.0, 0.5, 0.33, 0.25, 0.2 for $n = 1$ to 5. Calculate its spectral centroid.

Solution to Exercise 5.5

The spectral centroid is the amplitude-weighted mean frequency:

$$\text{centroid} = \frac{\sum A_n f_n}{\sum A_n}. \tag{5.17}$$

The numerator:

$$(1.0)(200) + (0.5)(400) + (0.33)(600) + (0.25)(800) + (0.2)(1000) \tag{5.18}$$

$$= 200 + 200 + 198 + 200 + 200 = 998. \tag{5.19}$$

The denominator:

$$1.0 + 0.5 + 0.33 + 0.25 + 0.2 = 2.28. \tag{5.20}$$

So

$$\text{centroid} = \frac{998}{2.28} = 438 \text{ Hz}. \tag{5.21}$$

Therefore, the centroid is about 438 Hz, roughly the 2nd harmonic. Note the pattern visible in the numerator: with amplitudes falling as $1/n$, every harmonic contributes equally to the centroid, so $1/n$ spectra sound as bright as they do.

Exercise 5.6

(Moderate) A spectrogram is computed with a window of 46 ms. (a) Approximately what frequency resolution does this give? (b) Could it separate the 1st and 2nd harmonics of a 100 Hz tone? (c) Could it resolve two events 10 ms apart in time?

Solution to Exercise 5.6

(a) The frequency resolution of a window of duration T is roughly $\Delta f \approx 1/T$:

$$\Delta f \approx \frac{1}{46 \times 10^{-3} \text{ s}} = 22 \text{ Hz}. \tag{5.22}$$

(b) Yes. The two harmonics are 100 Hz apart, comfortably more than the 22 Hz resolution.

(c) No. The window is 46 ms long, so two events 10 ms apart fall inside the same window and are smeared together.

Therefore, this window is good for frequency and poor for time. Resolving the 10 ms events would need a window under about 10 ms, which would degrade the frequency resolution to roughly 100 Hz, no longer able to separate the harmonics of part (b). This is the trade-off of §5.4, in numbers.

Exercise 5.7

(Straightforward) A clarinet's spectrum contains strong odd harmonics and weak even ones in its low register. Its lowest note sounds 147 Hz. (a) List the frequencies of its first four strong partials. (b) A flute at the same pitch has all harmonics. List its first four. (c) Explain the timbral consequence.

Solution to Exercise 5.7

(a) Odd harmonics of 147 Hz: 147, 441, 735, 1029 Hz.

(b) All harmonics: 147, 294, 441, 588 Hz.

(c) The clarinet is missing the octave (294 Hz), the second octave (588 Hz), and every other even partial. The octave above the fundamental is the partial that most reinforces a sense of a full, rounded tone, so removing it leaves the characteristic hollow, woody clarinet sound.

Therefore, the difference is not a matter of degree but of which partials exist at all: the same distinction the stopped and open pipes of Chapter 3 demonstrated.

Exercise 5.8

(Moderate) Two tones have identical harmonic amplitudes and different phases. (a) Do they have the same waveform? (b) Do they sound the same? (c) Do they carry the same energy? Justify each answer.

Solution to Exercise 5.8

(a) No. Phase determines how the harmonics line up, and shifting them changes the shape of the sum entirely, often changing the peak amplitude by a factor of two or more.

(b) Essentially yes, for a steady tone. This is Ohm's acoustic law: the ear analyzes into frequency components and is largely insensitive to the relative phases among them.

(c) Yes. The energy is $\sum A_n^2$, which contains no phases at all.

Therefore, the two differ in waveform, agree in energy, and agree in sound, which is only strange if one assumes the ear works on the waveform. It does not; it works on something much closer to the spectrum.

Exercise 5.9

(Straightforward) A guitar string is plucked 1/7 of the way along its length. (a) Which harmonics are missing? (b) Explain, in terms of this chapter's vocabulary, why guitar makers often place the bridge so that the string is *not* plucked at a simple fraction.

Solution to Exercise 5.9

(a) Plucking at $L/7$ puts the pluck at a node of the 7th mode and all its multiples, so harmonics 7, 14, 21, …are absent.

(b) Removing the 7th harmonic is, in this case, desirable rather than accidental. The 7th harmonic is 31 cents flat of any equal-tempered note (Chapter 9) and clashes with the harmonies of common-practice music, so plucking near $L/7$ is a way of suppressing a partial that would otherwise sound out of tune.

Therefore, the pluck point is a spectral design choice rather than an arbitrary one, and the same reasoning explains why piano hammers strike at about 1/7 of the string length.

Exercise 5.10

(Moderate) A sung vowel has formants at 700 Hz and 1200 Hz. It is sung first at $f_1 = 140$ Hz and then at $f_1 = 350$ Hz. (a) Which harmonic falls nearest each formant, at each pitch? (b) Comment on which pitch will render the vowel more clearly.

Solution to Exercise 5.10

(a) At $f_1 = 140$ Hz the harmonics are 140, 280, 420, 560, 700, 840, …The 5th harmonic (700 Hz) lands exactly on the first formant; the closest to 1200 Hz is the 9th at 1260 Hz, 60 Hz away.

At $f_1 = 350$ Hz the harmonics are 350, 700, 1050, 1400, …The 2nd (700 Hz) lands on the first formant; the closest to 1200 Hz is the 3rd at 1050 Hz or the 4th at 1400 Hz, both 150–200 Hz away.

(b) The low pitch. Its harmonics are spaced 140 Hz apart, so there is nearly always one close to any formant, and the formant peaks are well sampled. At the high pitch the harmonics are 350 Hz apart, and a formant can fall in a gap between them, leaving the ear with no evidence that it is there.

Therefore, the vowel is clearer at the lower pitch: the reason a soprano's words become hard to make out at the top of her range, and the reason singers modify vowels there (Chapter 13).

Exercise 5.11

(Challenging) Show that for a waveform built from harmonics with amplitudes A_n, the total energy is proportional to $\sum A_n^2$ and is independent of the phases.

Solution to Exercise 5.11

The energy of a signal over one period is proportional to the mean of its square:

$$\langle y^2 \rangle = \left\langle \left(\sum_n A_n \sin(2\pi n f_1 t + \phi_n) \right)^2 \right\rangle . \tag{5.23}$$

Expanding the square gives terms of two kinds. The "diagonal" terms are

$$A_n^2 \, \langle \sin^2(2\pi n f_1 t + \phi_n) \rangle = \tfrac{1}{2} A_n^2, \tag{5.24}$$

since the mean square of a sinusoid is $\frac{1}{2}$ whatever its phase. The "cross" terms, with $n \neq m$, are

$$A_n A_m \, \langle \sin(2\pi n f_1 t + \phi_n) \sin(2\pi m f_1 t + \phi_m) \rangle = 0, \tag{5.25}$$

by the orthogonality of §5.1, and this vanishes for *any* pair of phases.

So

$$\langle y^2 \rangle = \tfrac{1}{2} \sum_n A_n^2, \tag{5.26}$$

which contains no phases at all.

Therefore, the energy depends only on the amplitudes. This is Parseval's theorem, and it is the formal statement of why two signals with the same spectrum and different phases are equally loud however different they look.

Exercise 5.12

(Moderate) A note's partials decay with time constants $\tau_n = \tau_1 / n$: the nth partial decaying n times as fast as the fundamental. Take $\tau_1 = 2.0$ s. (a) After 1.0 s, what fraction of its initial amplitude does the 1st partial retain? The 6th? (b) Describe how the timbre changes over the note.

Solution to Exercise 5.12

(a) Using $A_n(t) = A_n(0)e^{-t/\tau_n}$ with $\tau_n = 2.0/n$:

$$\frac{A_1(1.0)}{A_1(0)} = e^{-1.0/2.0} = e^{-0.5} = 0.61, \tag{5.27}$$

$$\frac{A_6(1.0)}{A_6(0)} = e^{-1.0/(2.0/6)} = e^{-3.0} = 0.050. \tag{5.28}$$

(b) The high partials die far faster than the low ones, so the spectral centroid falls steadily and the note becomes progressively darker and more sine-like as it decays.

Therefore, the timbre of a plucked or struck note is not a fixed property but a moving one, bright at the onset and mellow at the end. This is exactly the behavior visible in the spectrogram of §5.4, and it is one of the things that makes a sustained synthesized tone sound artificial when it is left unchanged throughout.

Exercise 5.13

(Moderate) A synthesizer produces a tone with harmonics at exactly $1/n$ amplitude and no noise, no envelope variation, and no inharmonicity. Listeners describe it as "sterile". Name three specific additions that would make it more realistic, and say what each contributes.

Solution to Exercise 5.13

Three, drawing on this chapter:

1. **A transient onset**, partials arriving at slightly different times, with the high ones leading. §5.4 showed that the attack carries more identifying information than the steady spectrum, and a tone that simply switches on has none.

2. **Noise**, breath, bow scrape, key click, at the onset and continuing at low level. Noise has a continuous spectrum and so is entirely absent from a purely harmonic synthesis, yet every real instrument produces it.

3. **Time-varying partial amplitudes**, high harmonics decaying faster than low ones, and small independent fluctuations in each. A real note's spectrum is never constant, and Exercise 5.12 showed how much the timbre moves over a single note.

A fourth worth mentioning is slight inharmonicity or slight pitch instability, which smears each spectral line into a narrow band and removes the unnaturally perfect periodicity.

Therefore, what is missing from the sterile tone is precisely everything Fourier's theorem does not describe: the theorem covers steady periodic sound completely, and real instruments are never quite steady and never quite periodic.

Exercise 5.14

(Moderate) A square wave at 150 Hz is passed through a filter that removes everything above 1000 Hz. (a) Which harmonics survive? (b) Sketch in words how the waveform changes. (c) How does the sound change?

Solution to Exercise 5.14

(a) A square wave has odd harmonics only: 150, 450, 750, 1050, 1350, …Hz. Those below 1000 Hz are the 1st, 3rd, and 5th, 150, 450, and 750 Hz.

(b) With only three partials left, the sharp vertical edges disappear. What remains is a rounded, rippling approximation to a square, recognizably square-ish in outline, but with visible oscillation along the flat tops, since three harmonics is far too few to build a corner.

(c) The tone becomes much duller and less reedy. The pitch is unchanged at 150 Hz, because the fundamental and the harmonic spacing are untouched.

Therefore, the filter has changed the timbre substantially while leaving the pitch exactly where it was: a clean separation of the two attributes, and the basic operation of subtractive synthesis (Chapter 15).

Part III

Hearing and Perception

Chapter 6

The Ear and the Physiology of Hearing

Learning Objectives

By the end of this chapter, you should be able to:

- Name the principal structures of the outer, middle, and inner ear and state the function of each.
- Explain how the shape of the pinna and the resonance of the ear canal shape the sound reaching the eardrum, and relate the canal resonance to the region of greatest sensitivity.
- Explain the impedance-matching problem between air and the fluid of the cochlea, and calculate the pressure gain the middle ear achieves through the area ratio and the ossicular lever.
- Describe the structure of the cochlea and the traveling wave on the basilar membrane.
- State the place principle, and use the tonotopic map to predict where a given frequency produces its maximum response.
- Describe how hair cells transduce mechanical motion into neural signals, and distinguish the roles of inner and outer hair cells.
- Explain phase locking and the volley principle, and state the frequency range over which timing information is available to the auditory nerve.
- Explain how excessive sound exposure damages hearing, why the damage is permanent, and what the exposure limits are.
- Explain why hair-cell damage can produce tinnitus, a phantom sound, rather than simple silence at the affected frequency.

Introduction

The book now turns around. Chapters 1 to 5 followed sound from a vibrating object out into the room. The next three follow it from the room into a listener, and the story changes character completely.

Up to this point everything has been measurable with instruments. A microphone and an oscilloscope can tell you the frequency, the amplitude, and the spectrum of a sound, and two competent experimenters will agree on all three. From here on, the quantities are **judgments**: how loud something seems, how high it seems, whether two notes seem to go together. No instrument measures those. They are produced by a listener, and to understand them you have to understand the machinery that produces them.

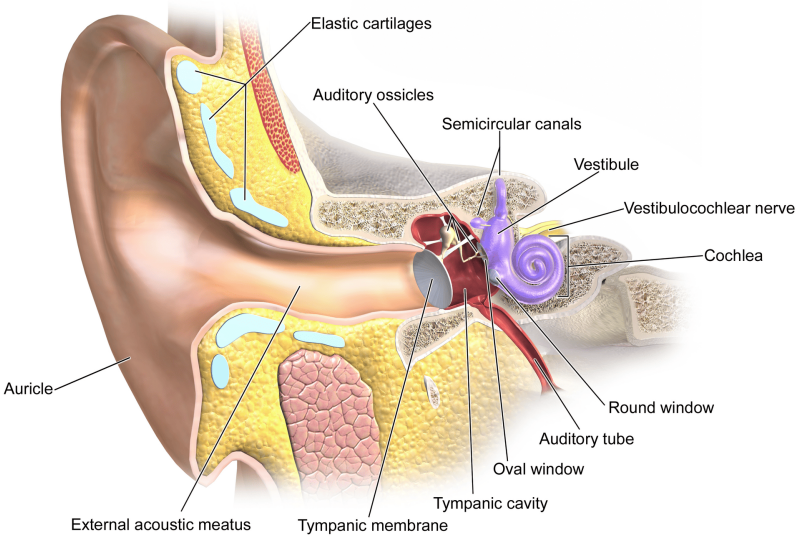


Figure 6.1: The ear is both anatomy and an acoustic pathway. This illustration gives the reader a physical object to keep in mind while the chapter develops its transduction model. BruceBlaus / Blausen Medical, CC BY 3.0.

That machinery begins with a mechanical problem that is worth appreciating before any biology appears. Sound in air must be transferred into fluid, and Chapter 2 showed that a boundary between media of very different acoustic impedance reflects almost everything. Air to water reflects about 99.9% of the incident energy: a loss of some 30 dB. An ear made simply of a membrane over a fluid-filled cavity would be nearly deaf.

The middle ear exists to solve exactly that problem, and it recovers almost all of the 30 dB. The rest of the chapter is about what the inner ear does with the sound once it has got it: how it separates frequencies, how it turns motion into nerve impulses, and what its limits are.

6.1 The Outer Ear

The Pinna and Direction

The visible ear, the **pinna**, is a collecting funnel, but that is the least interesting thing about it. Its folds and ridges are not decorative. They reflect incoming sound with small delays that depend on the direction it came from, and those delays reinforce some frequencies and cancel others.

The result is that the spectrum arriving at your eardrum is stamped with a direction-dependent pattern. A sound from above is filtered differently from one in front, which is differently from one behind.

This is the only cue available for distinguishing front from back, or up from down. The two-eared cues, a sound reaching one ear sooner and louder than the other, locate sound left and right very well, but they are identical for a source in front and the same source directly behind. The pinna breaks that ambiguity, and it is why a recording made with microphones

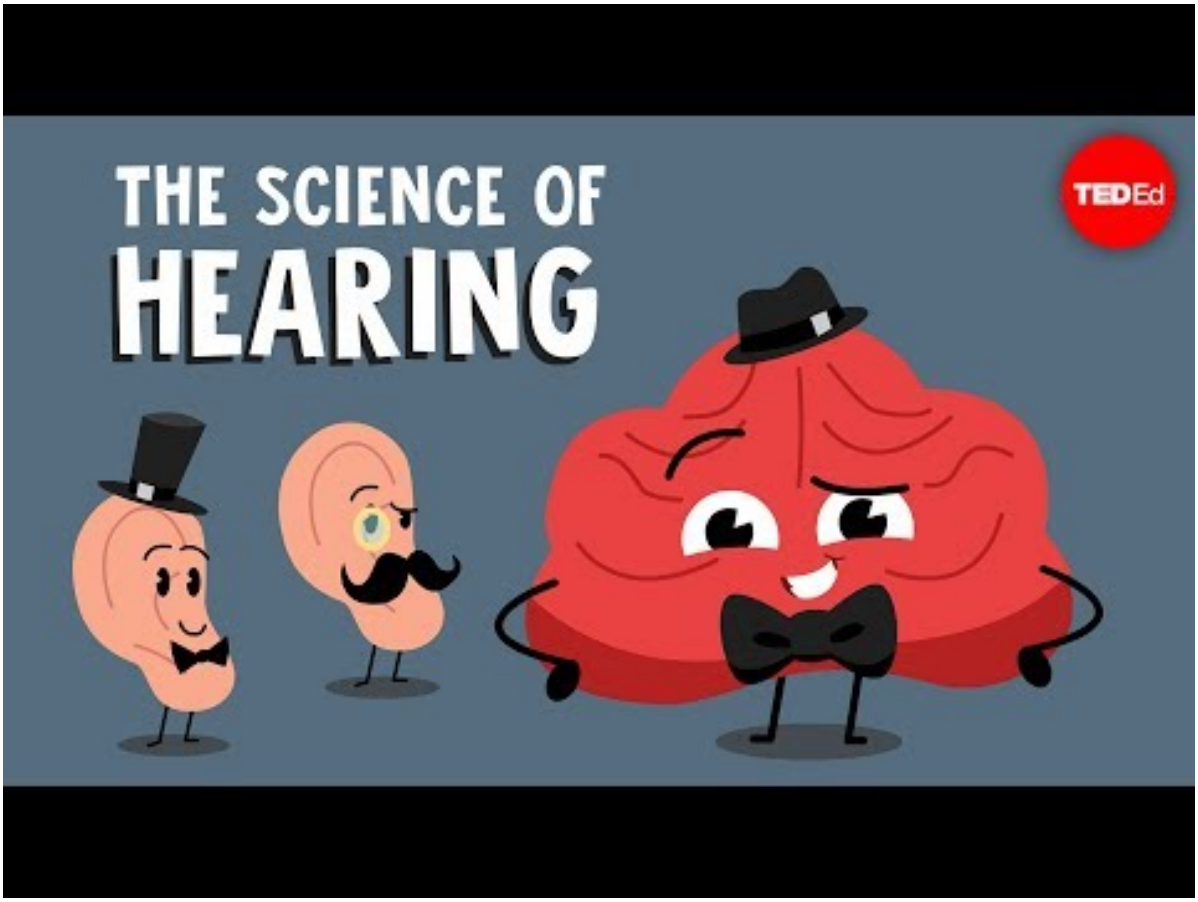


Figure 6.2: TED-Ed and Douglas L. Oliver give an anatomical overview of hearing, following a sound from the outer ear through the ossicles and cochlea. It provides a visual map for the structures examined separately below. Watch on YouTube: The Science of Hearing

in a dummy head's ears can produce a startlingly convincing sense of height and behindness through headphones (Chapter 15).

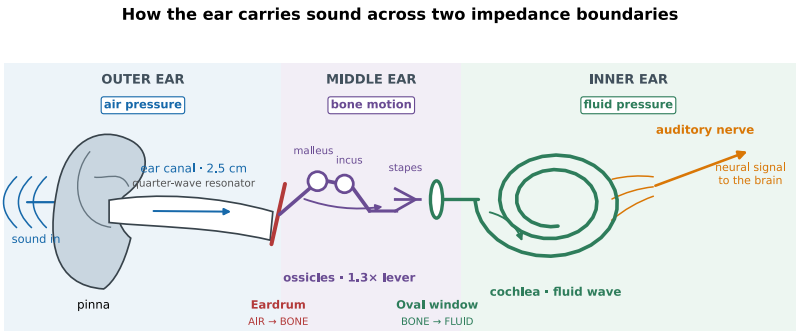


Figure 6.3: The ear, drawn for the argument rather than as an anatomy plate. Arrows follow one continuous signal path through air, bone, fluid, and finally the auditory nerve. The two changes of medium, at the eardrum and oval window, are where the interesting physics happens.

The Ear Canal as a Resonator

The ear canal is a tube about 25 mm long, closed at the far end by the eardrum. That is a stopped pipe, and Chapter 3 gives its fundamental resonance immediately:

$$f_1 = \frac{v}{4L} = \frac{343 \text{ m/s}}{4(0.025 \text{ m})} = 3.4 \text{ kHz}. \tag{6.1}$$

The canal therefore boosts sounds near 3–4 kHz by about 10 dB before they reach the eardrum at all.

This is not a coincidence, and it is a genuinely satisfying piece of physics. **The region of greatest hearing sensitivity is at 3–4 kHz**, and a good part of the reason is a quarter-wave resonance in a 2.5 cm tube. It is also the region of the consonants that carry most of the intelligibility of speech, and the region in which a singer's voice must compete with an orchestra (Chapter 13).

> **Self-check**
> Predict what happens to the ear canal's resonance when a person puts in an earplug that shortens the effective canal. Does the peak move up or down, and by roughly how much for a plug that occupies the outer third?

The Eardrum

The eardrum is a thin conical membrane, about 55 mm^2 in area, that moves when the pressure in front of it differs from the pressure behind. Its displacement at the threshold of hearing is astonishing: roughly 10^{-11} m, about a tenth the diameter of a hydrogen atom, and considerably less than the thermal jiggling of the membrane's own molecules.

That the ear works at all at such amplitudes is possible because it averages over time and area, and because it is not actually detecting a *displacement* below the thermal noise, it is detecting a *correlated* displacement against an uncorrelated background. Hearing sits very close to the physical limit of what a detector in a warm, noisy body could achieve.

6.2 The Middle Ear

Three Bones and a Lever

Behind the eardrum is an air-filled cavity containing the three smallest bones in the body: the **malleus**, **incus**, and **stapes**, or hammer, anvil, and stirrup. They form a chain from the eardrum to the **oval window**, the membrane-covered opening into the fluid-filled cochlea.

The chain is a lever. The hammer's arm is slightly longer than the anvil's, so the force at the stirrup is about 1.3 times the force at the eardrum, at 1.3 times less displacement.

Impedance Matching by Area Ratio

The lever is the smaller half of the trick. The larger half is area.

The eardrum has an area of about 55 mm^2; the oval window has an area of about 3.2 mm^2. The whole force collected over the eardrum is delivered to the much smaller oval window, and pressure is force over area, so the pressure is multiplied by the ratio of the areas, about 17.

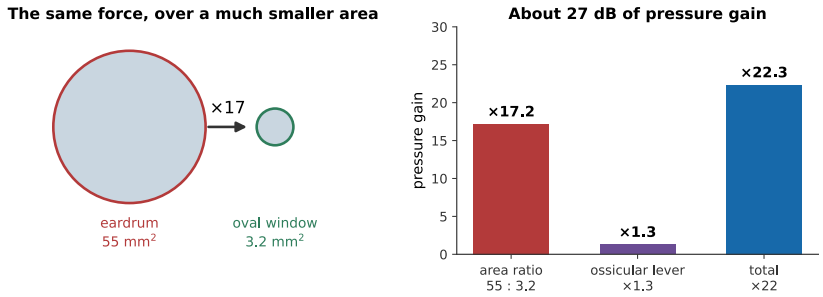


Figure 6.4: Where the middle ear's gain comes from. The area ratio does most of the work; the ossicular lever adds a little more. Together they multiply the pressure by about 22, which is 27 dB, close to the 30 dB that would otherwise be lost at the air–fluid boundary.

The total pressure gain is about $17 \times 1.3 \approx 22$, or 27 dB. Set against the 30 dB that a bare air–fluid boundary would lose, the middle ear recovers all but a few decibels of what would otherwise be thrown away.

Worked example: what hearing would be without a middle ear

Roughly 99.9% *of sound energy incident on a water surface from air is reflected. Express that loss in decibels, and compare with the middle ear's gain.*

The transmitted fraction is 0.001, so the level loss is

$$10 \log_{10}(0.001) = -30 \text{ dB}. \tag{6.2}$$

The middle ear supplies +27 dB. The net loss is therefore about 3 dB rather than 30 dB.

In terms of what that means for a listener: 27 dB of extra sensitivity is a factor of 500 in intensity. A conversation at four meters would become a conversation shouted from the next room. This is not a small refinement, it is most of the difference between hearing and not.

The Acoustic Reflex

The middle ear is not a passive lever. Two small muscles, the tensor tympani and the stapedius, contract in response to loud sound, stiffening the chain and reducing its transmission by up to 20 dB at low frequencies.

It is a protection mechanism, and it has two limitations worth knowing, because both bear on musicians. It takes 50–100 ms to engage, so it offers no protection at all against a gunshot or a snare hit. And it fatigues, so it does not protect against sustained loud sound either. The reflex is good at conversation-level fluctuations and useless against the two things that actually damage hearing.

6.3 The Cochlea

Anatomy of a Coiled Tube

The **cochlea** is a fluid-filled tube about 35 mm long, coiled into two and a half turns: the coiling is a packing convenience and has no acoustic significance. Running along its length is the **basilar membrane**, and the membrane's properties change systematically from one end to the other:

- At the **base**, near the oval window, it is narrow (about 0.1 mm) and stiff.
- At the **apex**, the far end, it is wide (about 0.5 mm) and floppy.

Narrow and stiff means a high natural frequency; wide and floppy means a low one, exactly as Chapter 1 would predict for anything on a spring. The membrane is therefore a continuous graded series of resonators, running from about 20 kHz at the base to about 20 Hz at the apex.

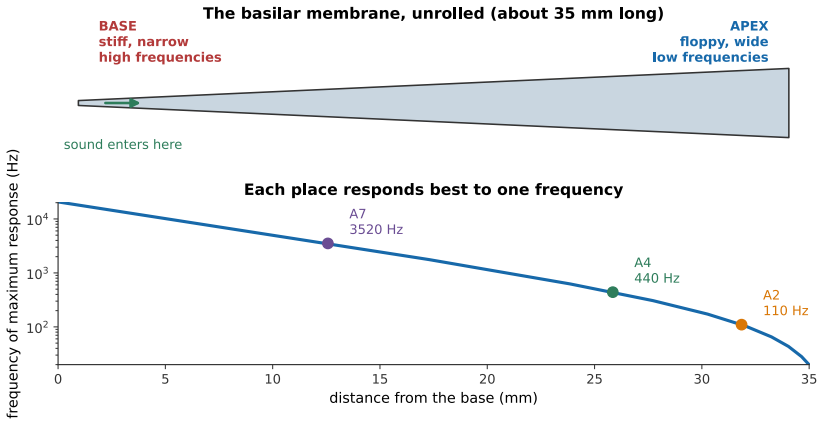


Figure 6.5: The basilar membrane unrolled, and the map of frequency onto place. The frequency axis is logarithmic and the distance axis is linear, and the relationship is close to a straight line, which means **equal musical intervals occupy equal distances along the cochlea**. An octave is about 3.5 mm, anywhere in the middle of the range.

That last observation is worth dwelling on. The cochlea is laid out logarithmically in frequency, which is the same way music is laid out. The octave is the same physical distance whether it is the octave above middle C or three octaves higher. Much of Chapter 8 follows from this one fact.

The Traveling Wave on the Basilar Membrane

Georg von Békésy won the 1961 Nobel Prize for working out what actually happens in there, and the answer is subtler than "each place resonates".

A sound entering at the oval window sets off a **traveling wave** that moves along the membrane from base toward apex. As it travels it grows, reaches a peak at the place matched to its frequency, and then dies away very rapidly.

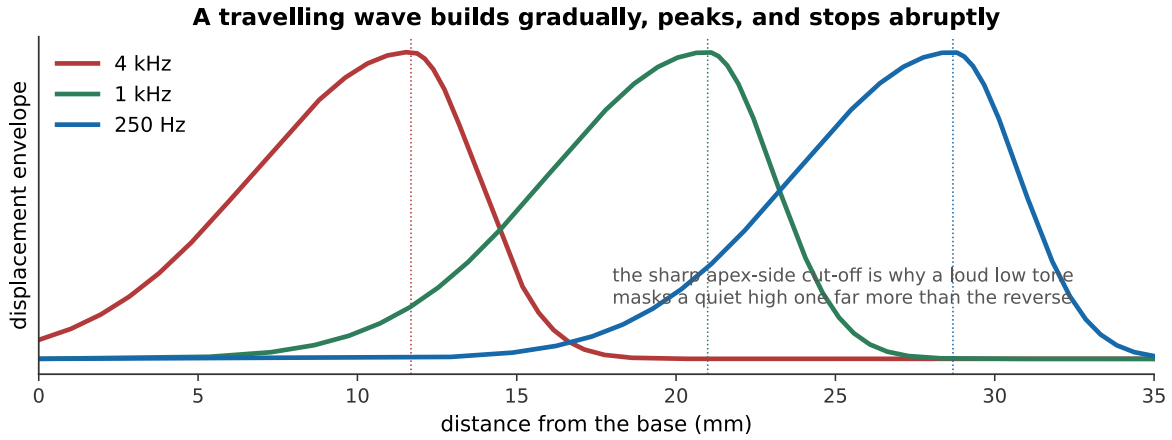


Figure 6.6: Envelopes of the traveling wave at three frequencies. Each rises gradually and falls steeply. The asymmetry is not cosmetic: a low-frequency wave travels the whole length of the membrane and excites everything on the way, while a high-frequency wave never reaches the apex at all.

The asymmetry of those envelopes has a direct musical consequence, and it is one of the few places where anatomy explains something a listener notices. **A loud low tone masks a quiet high one far more effectively than the reverse** (Chapter 7), because the low tone's traveling wave passes over, and excites, the region belonging to the high tone on its way to its own place. The high tone's wave never gets near the low tone's region at all. This is called **upward spread of masking**, and it is why a bass guitar buries a triangle and a triangle does not bury a bass guitar.

The Tonotopic Map

The place-to-frequency correspondence is called **tonotopic organization**, and it is preserved all the way up: the auditory nerve, the brainstem nuclei, and the auditory cortex are all laid out as maps of frequency. The brain receives something that is already sorted by frequency before any neural processing has happened.

This is the physiological basis of Ohm's acoustic law from Chapter 5. The ear is, mechanically, something very like a Fourier analyzer. And because it encodes *place*, while the relative timing between distant places is not preserved, the phase relationships among widely separated harmonics are discarded, exactly as the listening experiment showed.

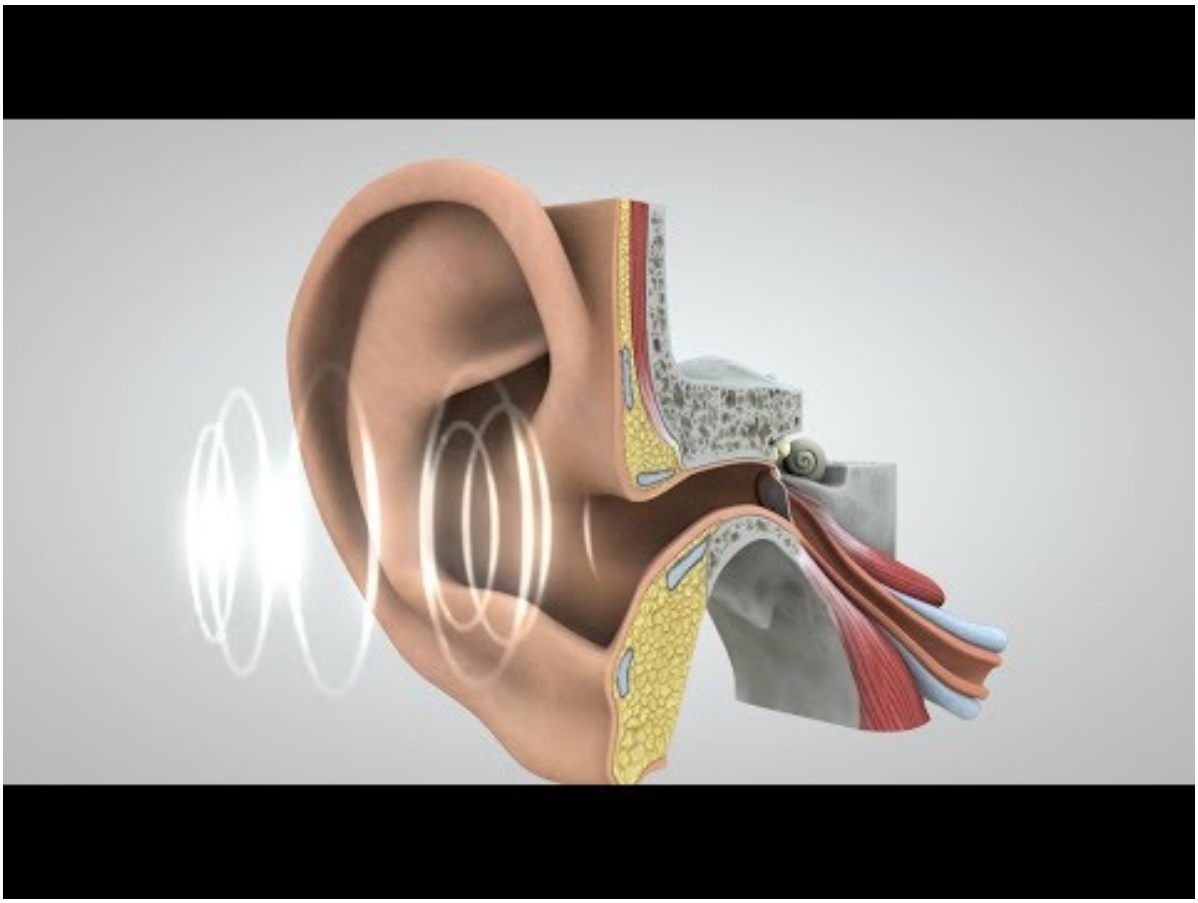


Figure 6.7: The National Institutes of Health follows vibration through the eardrum and ossicles, into the cochlea and hair cells, and finally along the auditory nerve. The animation connects the traveling wave on the basilar membrane to the neural signal it produces. Watch on YouTube: Journey of Sound to the Brain

6.4 From Motion to Nerve Impulse

Hair Cells and the Organ of Corti

Sitting on the basilar membrane is the **organ of Corti**, containing about 16,000 **hair cells**, each carrying a bundle of fine projections called stereocilia.

When the membrane moves, the stereocilia are bent sideways. Bending them one way pulls open molecular gates and lets ions flood into the cell; bending them the other way closes them. The cell's voltage therefore follows the motion of the membrane, and that voltage change triggers the release of neurotransmitter onto the auditory nerve fibers beneath.

The mechanism is direct, a mechanical gate, opened by a physical tug, so hearing is fast. There is no chemical cascade of the kind vision uses, and the latency is a matter of microseconds.

Inner and Outer Hair Cells

There are two kinds, in very different numbers, and their roles were a genuine surprise.

Inner hair cells, about 3,500 of them in a single row, are the sensors. They carry roughly 95% of the fibers going *to* the brain. These are what hearing is made of.

Outer hair cells, about 12,000 in three rows, are mostly not sensors at all. Most of their connections run *from* the brain, and their job is to **change length** in response to the voltage across them, fast enough to follow an audio-frequency signal.

That makes them an active amplifier. An outer hair cell detects the local motion and pushes in phase with it, feeding energy back into the membrane. The consequence is a sharpening of the traveling wave's peak, a rise in the effective Q of Chapter 4, by a factor of 100 to 1000 at low levels, and a corresponding improvement in both sensitivity and frequency selectivity.

> **Note**
> The cochlea's active amplifier can be caught in the act. A sensitive microphone in the ear canal can detect faint sounds coming *out* of a healthy ear, **otoacoustic emissions**, generated by the outer hair cells. Their presence is now the standard screening test for hearing in newborns, since it requires nothing of the infant at all. An ear that is not emitting is an ear whose amplifier is not working.

Outer hair cells are also the first casualties of noise exposure, so noise-induced hearing loss shows up first as a loss of *clarity* in noisy places rather than as a loss of volume: the amplifier has gone, taking the sharpening with it, and sounds that used to be separable now blur together.

Place, Timing, and the Volley Principle

There is a second stream of information alongside place.

Auditory nerve fibers tend to fire at a particular phase of the basilar membrane's motion, **phase locking**. A fiber does not fire on every cycle, but when it fires, it fires at the same point in the cycle. Pool many fibers and the aggregate firing pattern carries the waveform's period directly, a mechanism known as the **volley principle**.

Phase locking works up to about 4–5 kHz and degrades above that, because the nerve cannot maintain timing precision at those rates.

That number is suspicious. **The top note of a piano is 4186 Hz.** Almost all melodic instruments have their fundamentals below about 4 kHz: that is, within the range where timing information is available. Above it, pitch perception becomes markedly worse: listeners are poor at recognizing melodies made of pure tones above 5 kHz, and poor at tuning them. The musical range appears to stop about where the temporal mechanism does, and Chapter 8 takes up what that implies about how pitch is computed.



Figure 6.8: Veritasium demonstrates illusions that depend on how the ear and brain localize and interpret sound. They are reminders that hearing is an inference made from neural evidence, not a literal copy of the pressure waveform. Watch on YouTube: These Illusions Fool Almost Everyone

6.5 Limits and Damage

The Audible Range and How It Narrows

The nominal range of human hearing is 20 Hz to 20 kHz, and the upper figure applies to a young person with undamaged ears. It falls steadily with age, a process called **presbycusis**, at roughly a kilohertz per decade after the twenties, and it falls from the top down, because the base of the cochlea, which handles high frequencies, bears the brunt of a lifetime's exposure.

Noise-Induced Hearing Loss

The damage is mechanical, and it is permanent.

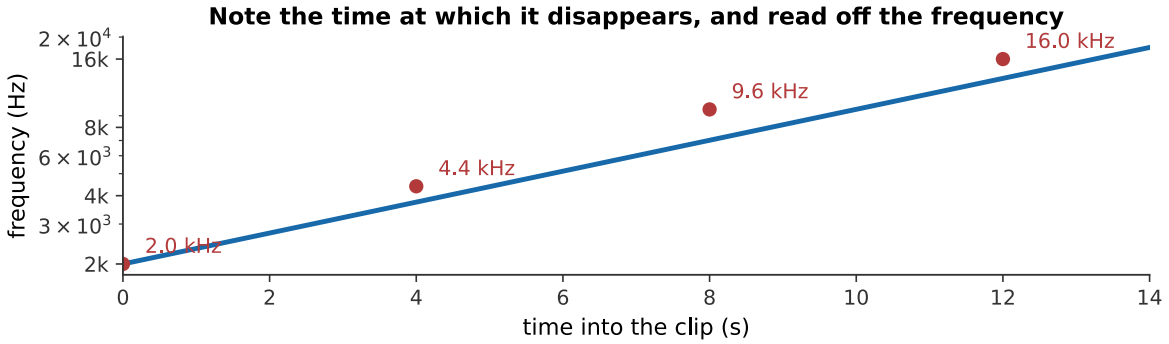


Figure 6.9: A sweep from 2 kHz to 18 kHz. Note the moment it disappears and read the frequency off the figure. Two warnings before drawing conclusions: many playback systems roll off above 15 kHz on their own, and the clip is *not* corrected for the ear's frequency response, so the apparent loudness changes as it climbs. This is an indication, not an audiogram. *What you hear:*A rising tone that climbs steadily for about fourteen seconds. At some point it becomes inaudible; when depends on the listener, on the playback system, and on the volume. Listen: Upper Limit Sweep

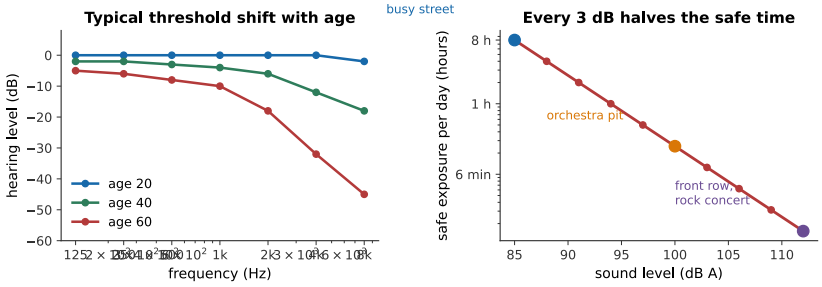


Figure 6.10: Left: typical threshold shift with age, which is concentrated at the high-frequency end. Right: safe exposure limits. Every 3 dB doubles the sound energy and therefore **halves** the time you can safely be exposed: the difference between 100 dB and 112 dB is the difference between fifteen minutes and under a minute.

Excessive sound overdrives the hair cells, and the stereocilia break or fuse. Hair cells in mammals **do not regenerate**. There is no recovery, no treatment, and no replacement, only prevention.

Three things make this more dangerous for musicians than for most people.

Temporary threshold shift is misleading. After a loud rehearsal, hearing is dulled and returns to normal within a day, which feels like the ear repairing itself. It is not. Each episode leaves a small permanent component, and the permanent components accumulate over years.

It is painless and gradual. Nothing hurts, and the loss appears first at frequencies above the range that carries most musical fundamentals. By the time a player notices, a great deal has usually been lost.

Damage can announce itself as a sound rather than a silence. Many musicians who lose hair cells at some frequency band do not experience that band as simply quieter; they experience **tinnitus**, a persistent ringing, hissing, or whining heard in the absence of any external sound. The leading explanation follows directly from the mechanism above: the brain's auditory pathway, deprived of its normal input at the damaged frequencies, turns up its own gain there, and the resulting spontaneous neural activity is heard as a tone. Tinnitus is common, often permanent once established, and, like the hearing loss that usually causes it, is easier to prevent than to treat.

The exposure limits are stricter than they look. The right-hand panel above is a 3 dB exchange rate: 85 dB for eight hours, 88 dB for four, 91 dB for two. An orchestra pit routinely reaches 95 dB, and a brass section measures over 110 dB at the player's own ear.

Exposure Limits for Musicians

The practical measures are unglamorous and they work.

Wear musician's earplugs. These are not foam plugs. Foam attenuates the high frequencies much more than the low ones, which makes music sound muffled and wrong, and players take them out. Musician's plugs use an acoustic filter to attenuate roughly evenly across the spectrum, typically 9, 15, or 25 dB, so the sound is quieter but not distorted. Nine decibels is a factor of eight in energy, and it multiplies safe exposure time by eight.

Use distance. The inverse-square law of Chapter 2 is free protection: doubling your distance from a loud source buys 6 dB, which quadruples the safe time.

Use time. Damage depends on total energy, so breaks are not merely a rest, they are a direct reduction in the dose.

> **Warning**
> This is the one part of the book with a practical consequence that cannot be undone. Hearing loss is cumulative, painless, and permanent, and the people most at risk from it are exactly the people most interested in this subject. Take the limits in the figure seriously.

6.6 Summary

- **The ear solves an impedance-matching problem.** An air–fluid boundary reflects about 99.9% of incident sound, a loss of 30 dB; the middle ear recovers about 27 dB of

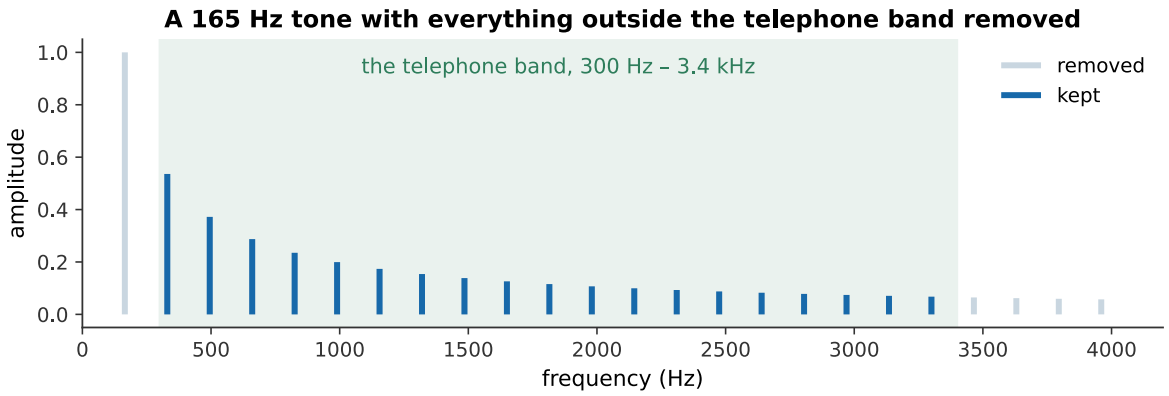


Figure 6.11: Everything outside 300 Hz to 3.4 kHz removed. Speech through this band stays perfectly intelligible, which is what the telephone was designed around, and it is roughly the band the outer and middle ear favor. Music through it does not survive nearly as well, which is a useful reminder that intelligibility and fidelity are different requirements. *What you hear:*The same note twice. The first is full and warm; the second is thin and boxy, like a voice on an old telephone, though the pitch is identical. Listen: Full spectrum, Telephone band only

it.

- **The pinna** imposes a direction-dependent filter on the spectrum, which is the only cue distinguishing front from back and up from down.
- **The ear canal is a stopped pipe** about 25 mm long, resonating near 3.4 kHz, which is a large part of why hearing is most sensitive there.
- **The middle ear's gain is mostly area ratio** (55 mm^2 to 3.2 mm^2, a factor of 17) with a smaller contribution from the ossicular lever (1.3), giving about 22 times the pressure, or 27 dB.
- **The basilar membrane is stiff and narrow at the base, floppy and wide at the apex**, so each place responds best to one frequency. The map is logarithmic: **equal musical intervals occupy equal distances**, an octave being about 3.5 mm.
- **A traveling wave** builds gradually and cuts off sharply past its peak. The asymmetry produces the **upward spread of masking**: low tones mask high ones far more than the reverse.
- **Inner hair cells sense; outer hair cells amplify.** The outer cells change length in phase with the motion, sharpening the response by a factor of 100–1000 and producing measurable otoacoustic emissions. They are also the first thing noise destroys.
- **Phase locking** carries timing information up to about 4–5 kHz, very nearly the top note of a piano, and roughly where reliable pitch perception stops.
- **Noise-induced hearing loss is mechanical, cumulative, painless, and permanent**, since mammalian hair cells do not regenerate. Every 3 dB halves the safe exposure time. Musician's earplugs, distance, and breaks all work.
- **Tinnitus**, a phantom ringing or hissing, often accompanies hair-cell damage: the brain raises its gain at frequencies it no longer receives input from, and the spontaneous activity is heard as sound.

6.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple response.** Which choices reduce risk during a loud rehearsal: flat-response musician's earplugs, quiet breaks, greater distance from loud sources, or relying on pain as an early warning?

2. **True or false.** Inner hair cells provide most cochlear amplification, while outer hair cells send most auditory information to the brain.

3. **Drag the words.** Complete the roles of the cochlear base and apex and of inner and outer hair cells.

4. **Fill in the blanks.** For the same risk, every increase of _____ dB cuts safe exposure time in _____.

5. **Mark the words.** Identify the structures responsible for directional filtering, impedance matching, and frequency separation: "The pinna supplies direction-dependent filtering, the middle ear matches air to cochlear fluid, and the basilar membrane separates frequencies by place."

6.8 Conceptual Questions

1. Explain why an ear consisting only of a membrane over a fluid-filled cavity would be about 30 dB less sensitive than a human ear, and what the middle ear does about it.

2. The ear canal is about 2.5 cm long. Explain why this produces a sensitivity peak near 3–4 kHz, and identify which boundary condition applies at each end.

3. Explain why the pinna is necessary for telling a sound in front from the same sound behind, given that the two ears alone locate sounds left and right perfectly well.

4. The basilar membrane's frequency map is logarithmic in frequency and linear in distance. State what this implies about the physical distance occupied by an octave, and connect it to why the octave is musically special.

5. A loud bass note masks a quiet high note much more than a quiet high note masks a loud bass note. Explain this using the shape of the traveling wave.

6. Distinguish the roles of inner and outer hair cells, and explain why damage to the outer cells shows up first as difficulty understanding speech in a noisy room rather than as a loss of loudness.

7. The acoustic reflex takes about 50 ms to engage and fatigues over time. Explain why this makes it useless against the two kinds of exposure that actually damage hearing.

8. Phase locking fails above about 4–5 kHz, and the highest note on a piano is 4186 Hz. Discuss whether this is likely to be a coincidence.

9. Tinnitus is usually described as a sound heard in the absence of any external stimulus, yet it typically follows damage at a *specific* frequency band. Explain why the pitch of the phantom sound and the location of the damage are connected, using the tonotopic map.

6.9 Problems

Exercise 6.1
(Straightforward) The ear canal is 2.5 cm long and behaves as a stopped pipe. Take $v = 343$ m/s. (a) Find its fundamental resonance. (b) Find its next resonance. (c) Is the second resonance inside the range of hearing?

Solution to Exercise 6.1
(a) A stopped pipe has $f_n = nv/4L$ for odd n:

$$f_1 = \frac{343 \text{ m/s}}{4(0.025 \text{ m})} = 3430 \text{ Hz}. \tag{6.3}$$

(b) The next mode is the third harmonic:

$$f_3 = 3(3430) = 10.3 \text{ kHz}. \tag{6.4}$$

(c) Yes, comfortably, 10.3 kHz is well within the 20 Hz to 20 kHz range, though the ear is considerably less sensitive there than at 3.4 kHz.

Therefore, the canal contributes resonances at about 3.4 kHz and 10 kHz, and the first of these coincides with the region of greatest sensitivity.

Exercise 6.2
(Moderate) The eardrum has area 55 mm^2 and the oval window 3.2 mm^2. The ossicular lever gives a force advantage of 1.3. (a) Find the total pressure gain. (b) Express it in decibels. (c) What would the gain be if the lever were absent?

Solution to Exercise 6.2
(a) Pressure gain is the area ratio times the lever ratio:

$$G = \frac{55 \text{ mm}^2}{3.2 \text{ mm}^2} \times 1.3 = 17.2 \times 1.3 = 22.3. \tag{6.5}$$

(b) For a pressure ratio, in decibels:

$$20 \log_{10}(22.3) = 27.0 \text{ dB}. \tag{6.6}$$

(c) Without the lever, $G = 17.2$, which is $20 \log_{10}(17.2) = 24.7$ dB.

Therefore, the lever contributes only 2.3 dB of the 27 dB total. The area ratio does almost all of the work.

Exercise 6.3
(Moderate) The threshold of hearing at 1 kHz corresponds to an eardrum displacement of about 1×10^{-11} m. (a) Compare this with the diameter of a hydrogen atom, about 1×10^{-10} m. (b) At the threshold of pain, 120 dB above threshold, what is the displacement?

Solution to Exercise 6.3

(a) The ratio is

$$\frac{1 \times 10^{-11}}{1 \times 10^{-10}} = 0.1, \tag{6.7}$$

so the threshold displacement is about a tenth of an atomic diameter.

(b) A level difference of 120 dB in pressure corresponds to an amplitude ratio of

$$10^{120/20} = 10^{6}. \tag{6.8}$$

So the displacement at the threshold of pain is

$$(1 \times 10^{-11}\ \text{m})(10^{6}) = 1 \times 10^{-5}\ \text{m} = 10\ \mu\text{m}. \tag{6.9}$$

Therefore, the eardrum's working range runs from a tenth of an atomic diameter to about a hundredth of a millimeter: a millionfold span, handled by one membrane.

Exercise 6.4

(Straightforward) The basilar membrane is 35 mm long and maps roughly logarithmically onto 20 Hz to 20 kHz. (a) How many octaves does it span? (b) How many millimeters does one octave occupy? (c) A semitone?

Solution to Exercise 6.4

(a) The range is a factor of 1000, so

$$n = \log_2(1000) = 9.97 \approx 10 \text{ octaves}. \tag{6.10}$$

(b) Dividing the length by the octaves:

$$\frac{35\ \text{mm}}{9.97} = 3.5\ \text{mm per octave}. \tag{6.11}$$

(c) One semitone is a twelfth of an octave:

$$\frac{3.5\ \text{mm}}{12} = 0.29\ \text{mm}. \tag{6.12}$$

Therefore, a semitone occupies about three tenths of a millimeter of membrane. Given that the just-noticeable frequency difference is far smaller than a semitone (Chapter 8), the ear plainly resolves pitch far more finely than this spacing alone would allow, which is one of several reasons place cannot be the whole story.

Exercise 6.5

(Moderate) A musician is exposed to 94 dB for an eight-hour rehearsal day. (a) Using a 3 dB exchange rate with a limit of 85 dB for eight hours, what is the permitted daily exposure at 94 dB? (b) By how many times has the musician exceeded it? (c) What attenuation of earplug would bring them inside the limit?

Solution to Exercise 6.5

(a) Each 3 dB halves the allowed time. From 85 to 94 dB is 9 dB, or three halvings:

$$T = \frac{8\ \text{h}}{2^3} = 1\ \text{h}. \tag{6.13}$$

(b) Eight hours against one hour permitted: a factor of 8.

(c) To come back inside the limit the level must drop by 9 dB, so a plug of 9 dB attenuation or more: the lowest rating musician's plugs are commonly made in.

Therefore, a 94 dB rehearsal day is eight times the safe dose, and the standard 9 dB musician's plug is exactly enough to fix it.

Exercise 6.6

(Moderate) A violinist measures 103 dB at their left ear and 91 dB at their right. (a) Find the permitted daily exposure for each ear. (b) Explain why the asymmetry arises and what practical step would reduce it.

Solution to Exercise 6.6

(a) Using the 3 dB rule from 85 dB for 8 h:

Left ear, $103 - 85 = 18$ dB, six halvings:

$$T = \frac{8}{2^6} = 0.125\ \text{h} = 7.5\ \text{minutes}. \tag{6.14}$$

Right ear, $91 - 85 = 6$ dB, two halvings:

$$T = \frac{8}{2^2} = 2\ \text{h}. \tag{6.15}$$

(b) The left ear is very close to the instrument, perhaps 15 cm from the f-holes, while the right is nearly an arm's length away and shadowed by the head. The inverse-square law of Chapter 2 accounts for most of the 12 dB.

The practical step is an earplug in the left ear at minimum. Left-ear hearing loss is a well-documented occupational pattern among violinists and violists, and it follows directly from this geometry.

Exercise 6.7

(Straightforward) A sound at 250 Hz produces its peak response about 26 mm from the base of the cochlea, and a sound at 4 kHz about 12 mm from the base. (a) How far apart are the two peaks? (b) Using 3.5 mm per octave, check this against the musical interval between the two frequencies.

Solution to Exercise 6.7

(a) $26 - 12 = 14$ mm.

(b) The frequency ratio is $4000/250 = 16$, which is

$$\log_2(16) = 4\ \text{octaves}. \tag{6.16}$$

At 3.5 mm per octave that predicts

$$4 \times 3.5 = 14\ \text{mm}. \tag{6.17}$$

Therefore, the two agree exactly, confirming that the map really is logarithmic: four octaves is four octaves' worth of membrane wherever on the cochlea it falls.

Exercise 6.8
(Moderate) The outer hair cells sharpen the cochlear response by raising its effective Q from about 1 to about 10 at low levels. Using the definitions of Chapter 4: (a) By what factor does the bandwidth at 1 kHz narrow? (b) Express each bandwidth in semitones.

Solution to Exercise 6.8
(a) Bandwidth is $\Delta f = f_0/Q$, so raising Q from 1 to 10 narrows the bandwidth by a factor of 10.

At 1 kHz: from $\Delta f = 1000$ Hz to $\Delta f = 100$ Hz.

(b) A bandwidth of 1000 Hz centered on 1000 Hz runs roughly 500 to 1500 Hz:

$$12 \log_2(1500/500) = 12(1.585) = 19 \text{ semitones}. \tag{6.18}$$

A bandwidth of 100 Hz runs roughly 950 to 1050 Hz:

$$12 \log_2(1050/950) = 12(0.1444) = 1.7 \text{ semitones}. \tag{6.19}$$

Therefore, the active amplifier takes the ear's frequency resolution from about an octave and a half down to under two semitones. Losing the outer hair cells does not merely make sounds quieter; it makes them run together, which is exactly what people with noise damage report.

Exercise 6.9
(Straightforward) A person's hearing threshold at 4 kHz has shifted by 35 dB. (a) By what factor in intensity has their sensitivity fallen? (b) A cymbal produces 4 kHz energy at 70 dB SPL at their seat. Do they hear it?

Solution to Exercise 6.9
(a) A 35 dB shift in intensity terms is

$$10^{35/10} = 3.2 \times 10^3, \tag{6.20}$$

so the sound must be about 3200 times more intense to be detected.

(b) Their threshold at 4 kHz is now 35 dB SPL rather than roughly 0. A 70 dB SPL cymbal is 35 dB above their threshold, so yes, they hear it, but at what a normal listener would experience as a much quieter sound.

Therefore, moderate high-frequency loss does not silence the cymbal; it dulls it. Such loss often goes unnoticed for a long time, and is usually detected first by an audiogram rather than by the person.

Exercise 6.10
(Moderate) Estimate the cochlea's "channel count" by treating it as a bank of filters each one critical band wide. Take the critical band as roughly a third of an octave across most of the range, and the range as 10 octaves. (a) How many filters is that? (b) Compare with the number of inner hair cells, about 3500.

Solution to Exercise 6.10

(a) At three filters per octave over ten octaves:

$$N = 3 \times 10 = 30 \text{ filters.} \tag{6.21}$$

(b) There are about 3500 inner hair cells, so roughly

$$\frac{3500}{30} \approx 120 \tag{6.22}$$

inner hair cells per critical band.

Therefore, the cochlea has of order thirty independent frequency channels but more than a hundred sensors in each. The redundancy is what allows the system to average away noise, which is how a detector working below the thermal-motion limit can function at all, and it also explains why substantial hair-cell loss can occur before any change is noticed.

Exercise 6.11

(Moderate) Phase locking in the auditory nerve fails above about 4.5 kHz. (a) What is the period of a 4.5 kHz tone? (b) A nerve fiber's refractory period, the minimum time between firings, is about 1 ms. How many cycles of a 4.5 kHz tone pass during it? (c) Explain how the volley principle allows the population to encode a period no single fiber can follow.

Solution to Exercise 6.11

(a) $T = 1/4500 = 2.2 \times 10^{-4}$ s, or 0.22 ms.

(b) 1 ms/0.22 ms = 4.5 cycles.

(c) No single fiber can fire on every cycle: it must sit out at least four of every five. But each fiber, when it does fire, fires at the same *phase*. Different fibers sit out different cycles, so pooling across many fibers reconstructs a firing pattern with a peak on every cycle, even though no individual contributor managed one.

Therefore, the population encodes a periodicity far faster than any of its members. This is the volley principle, and it is the mechanism that gives the auditory system timing information across the whole musically useful range.

Exercise 6.12

(Moderate) A musician can either (a) move from 2 m to 4 m from a loud source, or (b) wear a 9 dB earplug. Which gives more protection, and by how much? Assume the source radiates freely.

Solution to Exercise 6.12

(a) Doubling the distance in a free field reduces the level by

$$20 \log_{10}(2) = 6.0 \text{ dB.} \tag{6.23}$$

(b) The earplug gives 9 dB.

The earplug wins by 3 dB, which, on the 3 dB exchange rate, is a further factor of two in safe exposure time.

Therefore, the plug is better, and the two combine: moving *and* plugging gives 15 dB, which is a factor of 32 in permitted time. It is also worth noting that in a real hall the distance benefit would be less than 6 dB, because beyond the critical distance of Chapter 2 the reverberant field stops falling off with distance at all.

Exercise 6.13
(Challenging) Explain quantitatively why a foam earplug is a poor choice for a musician, given that it typically attenuates 10 dB at 250 Hz and 35 dB at 4 kHz, while a musician's plug attenuates about 15 dB at both.

Solution to Exercise 6.13
The foam plug's attenuation differs by

$$35 - 10 = 25 \text{ dB} \tag{6.24}$$

between 250 Hz and 4 kHz. In amplitude terms the high frequencies are attenuated

$$10^{25/20} = 18 \tag{6.25}$$

times more than the low ones.

The effect is a drastic tilt of the spectrum: an instrument's fundamental passes nearly unaffected while its upper harmonics are cut to a twentieth. Since Chapter 5 showed that the upper harmonics are what timbre is made of, the result is a muffled, boxy sound in which a player can neither judge their own tone nor hear the rest of the ensemble properly.

The musician's plug attenuates by the same 15 dB at both ends, so the spectrum is preserved and only the level changes.

Therefore, the foam plug offers more attenuation but destroys the information the musician needs, so it gets taken out, and a plug that is not worn attenuates nothing at all.

Exercise 6.14
(Challenging) A tone at 200 Hz and a tone at 3000 Hz are played at the same level. (a) Using the traveling-wave picture, explain which is more effective at masking the other. (b) A recording engineer complains that a bass guitar is "burying" a hi-hat. Suggest, from this chapter's physics, why reducing the bass level helps more than raising the hi-hat.

Solution to Exercise 6.14
(a) The 200 Hz tone. Its traveling wave must run almost the full length of the basilar membrane to reach its own place near the apex, and it excites the 3000 Hz region on the way past. The 3000 Hz wave peaks near the base and dies before reaching the apex, so it never disturbs the low-frequency region at all.

(b) Raising the hi-hat increases the total level without changing the fact that the bass is still spreading excitation across the hi-hat's region, and a louder mix simply moves both closer to the point where masking is strongest. Reducing the bass removes the masker itself, which lets the hi-hat through at a level it was already loud enough to reach.

Therefore, masking is asymmetric, and the remedy is to attend to the masker rather than the masked. Mixing engineers talk about "making space" by cutting rather than boosting for exactly this reason: it is the asymmetry the traveling wave predicts.

Chapter 7

Loudness, Decibels, and Equal-Loudness Contours

Learning Objectives

By the end of this chapter, you should be able to:

- Distinguish sound pressure, sound intensity, and loudness, and state which are physical and which is perceptual.
- Explain why the enormous range of audible intensities makes a logarithmic scale the natural one.
- Calculate sound intensity level and sound pressure level in decibels from intensity or pressure, and convert back.
- Explain why a factor of ten in intensity is 10 dB and a factor of two is 3 dB, and use those two facts to do decibel arithmetic without a calculator.
- Combine the levels of two or more incoherent sources, and explain why doubling the number of violins does not double the loudness.
- Distinguish the decibel as a ratio from the decibel as an absolute level referred to a stated reference, and use dB SPL correctly.
- Read the equal-loudness contours, define the phon, and explain why bass disappears first when music is played quietly.
- Define the sone, use the rule that ten phons doubles the loudness, and explain why a 60-piece orchestra is not sixty times as loud as one instrument.
- Explain masking and the critical band, and give a musical consequence of each.

Introduction

Turn the volume knob halfway down. Is the music half as loud?

Nobody thinks so, and the reason is the subject of this chapter. Between the sound pressure at your eardrum and the loudness you experience lie two transformations, and neither is a simple proportion.

The first is **enormous range**. The quietest audible sound and the loudest tolerable one differ in pressure by a factor of about a million, and in energy by a factor of about a million million. No linear scale copes with that, and the decibel exists to tame it.

The second is **frequency dependence**. The ear is far more sensitive at some frequencies than at others, and, the awkward part, *how much* more depends on how loud the sound is. A bass note and a mid-range note that match in loudness when played loudly will not match



Figure 7.1: Sound level is always measured somewhere. A concert hall makes that "somewhere" consequential: source, distance, reflections, and listener all enter the result. Jordan Fischer, CC BY 2.0.

when played quietly. Music therefore sounds thin at low volume, and it is a fact about you rather than about your loudspeakers.

This chapter separates three things that ordinary speech runs together:

Quantity	What it is	Units
Sound pressure level	Physical, measurable with an instrument	dB SPL
Loudness level	How loud it sounds, expressed as the level of an equally loud 1 kHz tone	phon
Loudness	How loud it sounds, on a scale where twice the number means twice as loud	sone

Only the first is physics in the sense of Chapters 1 to 5. The other two are properties of a listener, and getting from the first to the third is what the chapter does.

7.1 Pressure, Intensity, and the Range of Hearing

Two Physical Measures

Two physical quantities describe the strength of a sound, and both are used.

Sound pressure p is the amplitude of the pressure variation, in pascals. It is what a microphone and an eardrum actually respond to.

Intensity I is the power per unit area carried by the wave, in watts per square meter. It is related to pressure by

$$I = \frac{p^2}{\rho v}, \tag{7.1}$$

where ρv is the acoustic impedance of the medium, about 413 kg m^{-2} s^{-1} for air at room temperature.

The essential relation is the one Chapter 1 established: **intensity goes as the square of pressure.** Double the pressure and you quadruple the intensity. Every factor-of-two confusion in acoustics comes from losing track of this.

Twelve Orders of Magnitude

The numbers are worth writing out.

	Pressure (Pa)	Intensity (W/m^2)
Threshold of hearing at 1 kHz	2×10^{-5}	1×10^{-12}
Quiet conversation	2×10^{-3}	1×10^{-8}
Orchestra, fortissimo	2×10^{-1}	1×10^{-4}
Threshold of pain	20	1

The pressure range is 10^6 and the intensity range is 10^{12}. For comparison, the total acoustic power radiated by a full symphony orchestra playing as loudly as it can is somewhere around one watt, less than a small light bulb. The ear is an extraordinarily sensitive instrument working with extraordinarily small quantities.

A linear scale is hopeless here. If one millimeter on a graph represented the threshold of hearing, the threshold of pain would be a thousand kilometers away.

The Reference Values

The values in the first row of that table are not accidents; they are **definitions**, chosen because they are close to the average threshold of hearing at 1 kHz:

$$p_0 = 2 \times 10^{-5}\ \text{Pa} = 20\ \mu\text{Pa}, \qquad I_0 = 1 \times 10^{-12}\ \text{W/m}^2. \tag{7.2}$$

Every decibel figure in this book is referred to one of these, and they are consistent with each other: a pressure of p_0 in air carries an intensity of very nearly I_0.

7.2 The Decibel

A Logarithmic Scale

The **decibel** compresses a huge range into a manageable one by taking a logarithm. For a ratio of intensities,

$$L = 10 \log_{10}\left(\frac{I}{I_0}\right)\ \text{dB}. \tag{7.3}$$

Because intensity goes as pressure squared, and $\log(p^2) = 2\log(p)$, the same level in terms of pressure is

$$L = 20 \log_{10}\left(\frac{p}{p_0}\right) \text{ dB}. \tag{7.4}$$

The factor is 10 for power quantities and 20 for amplitude quantities. They give the same answer for the same physical sound; the difference is entirely in which quantity you started from. Using 10 where you need 20 halves every answer, and it is the most common error in the subject.

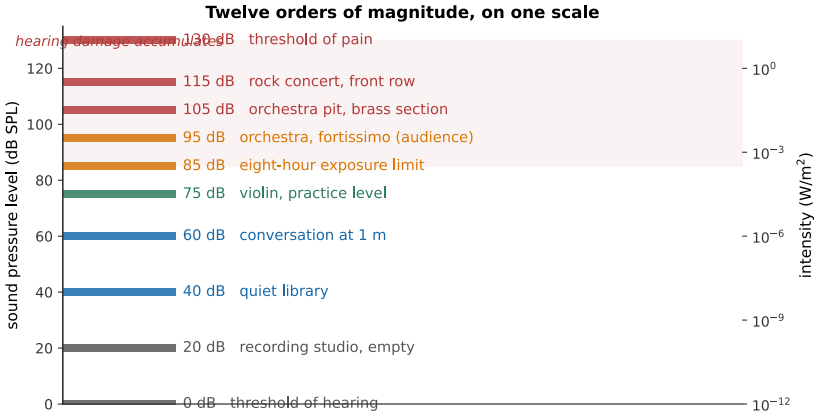


Figure 7.2: The decibel scale, with landmarks. The right-hand axis shows the intensities the same sounds correspond to: each 30 dB is a factor of a thousand. Note how compressed the *audible* world is on a decibel scale and how spread out it is in watts.

dB SPL and Other References

A decibel on its own is a **ratio** and says nothing absolute. "The passage is 6 dB louder" is meaningful. "The passage is 6 dB" is not, until a reference is stated.

dB SPL means referred to $p_0 = 20$ μPa, and is the one used throughout this book for absolute levels. Others exist, dBu and dBV in audio electronics, dBFS in digital audio (Chapter 15), and mixing them up produces errors of tens of decibels.

You will also meet **dB A**, which is a measurement weighted by a filter approximating the ear's sensitivity at moderate levels. Exposure limits are quoted in dB A precisely because a sound's damage potential depends on where its energy lies, not just how much there is.

Decibel Arithmetic Without a Calculator

Two facts make most decibel problems mental arithmetic.

$$\text{a factor of 10 in intensity} = 10 \text{ dB}, \qquad \text{a factor of 2 in intensity} = 3 \text{ dB}. \tag{7.5}$$

The second is approximate, $10 \log_{10} 2 = 3.01$, and the approximation is excellent.

From those two, by adding:

Intensity ratio	dB	Amplitude ratio	dB
2	3	2	6
4	6	4	12
10	10	10	20
100	20	100	40
1000	30	1000	60

Self-check
A sound is 23 dB above another in intensity. Without a calculator, what is the intensity ratio? (Break 23 into tens and threes.)

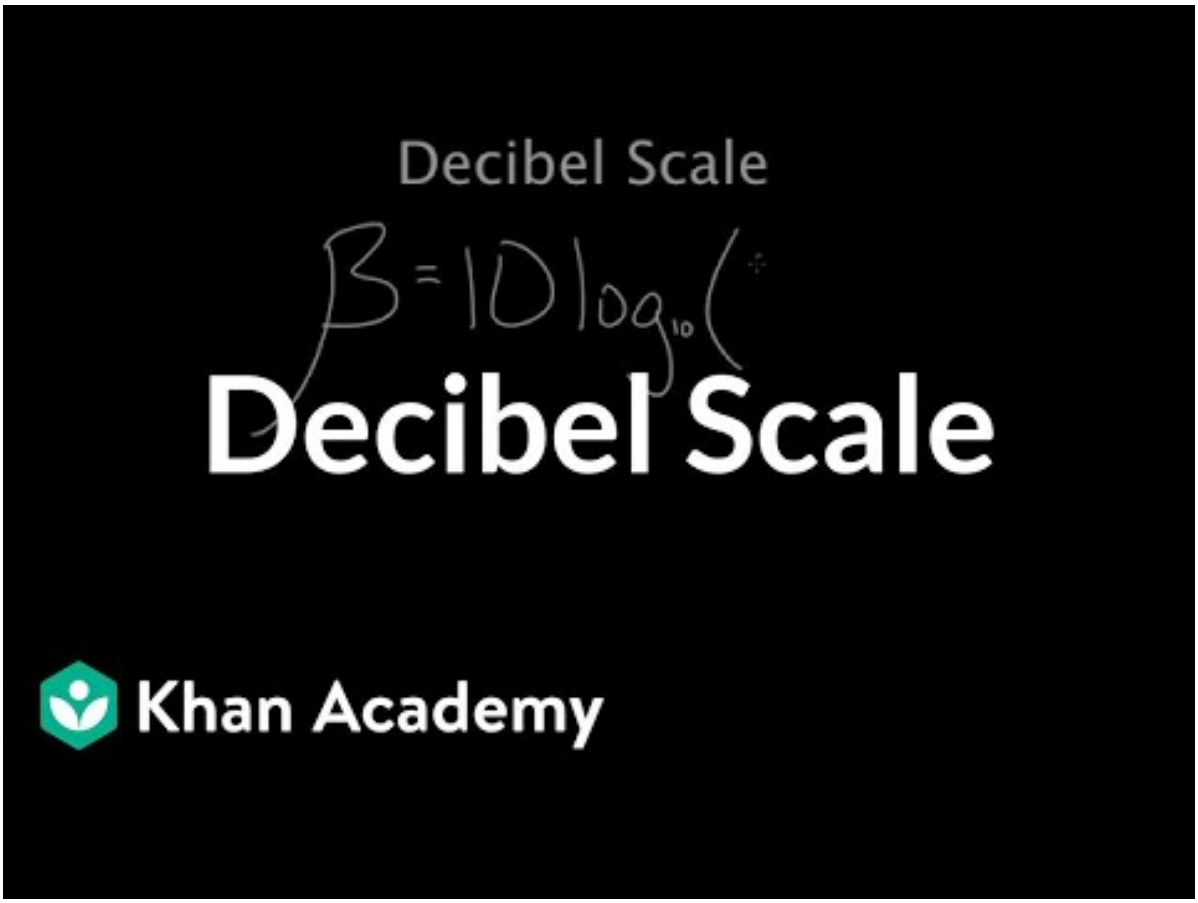


Figure 7.3: Khan Academy walks through the logarithmic decibel scale and the meaning of a 10 dB change. The presentation reinforces why multiplication of physical intensity becomes addition on a level scale. Watch on YouTube: Decibel Scale

7.3 Combining and Comparing Levels

Adding Incoherent Sources

Two violinists play the same note. How much louder than one?

The answer depends on whether their sounds are **coherent**, locked in a fixed phase relationship, or not.

Two loudspeakers fed from the same signal are coherent, and their pressures add: $2p$, which is 6 dB. Two violinists are *not* coherent: their frequencies and phases drift independently, so the interference at any point averages away over time and their **intensities** add instead: $2I$, which is 3 dB.

Real musicians are incoherent, so for n identical independent sources,

$$L_n = L_1 + 10\log_{10}(n). \tag{7.6}$$

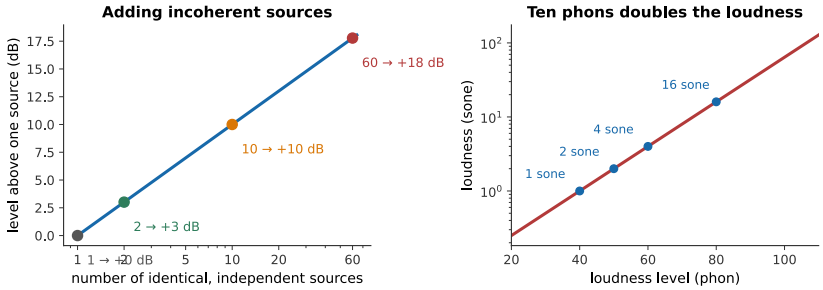


Figure 7.4: Left: adding incoherent sources. Sixty violins are only 18 dB louder than one. Right: loudness in sones against loudness level in phons, each 10 phons doubles the loudness.

The consequence for orchestration is sharp and slightly deflating. Sixty violins produce $10\log_{10}(60) = 18$ dB more than one. And since §7.5 will show that 10 phons is a doubling of loudness, 18 dB is under two doublings: **sixty violins sound less than four times as loud as one.** Orchestras have so many strings and so few trumpets for that reason: the section needs numbers merely to keep up.

Distance, Level, and the Six-Decibel Rule

Combining the inverse-square law of Chapter 2 with the definition of the decibel:

$$L_2 - L_1 = 20\log_{10}\left(\frac{r_1}{r_2}\right). \tag{7.7}$$

Doubling the distance therefore costs $20\log_{10}(2) = 6$ dB. Ten times the distance costs 20 dB.

This holds outdoors. Indoors it fails beyond the critical distance, for the reason Chapter 2 gave and Chapter 14 develops: most of what reaches a distant listener has bounced.

Worked example: a hall's back row

A soloist produces 85 dB at 1 m. What level reaches a listener 32 m away, (a) outdoors and (b) in a hall where the reverberant field holds the level constant beyond a critical distance of 6 m?

(a) Outdoors, 32 m is five doublings from 1 m:

$$L = 85 - 5(6) = 55 \text{ dB}. \tag{7.8}$$

(b) Indoors, the level falls by the inverse-square law only out to 6 m. From 1 m to 6 m:

$$20 \log_{10}(6) = 15.6 \text{ dB}, \tag{7.9}$$

so the level at the critical distance is about 69 dB, and it stays near that out to 32 m.

The difference between the two answers is 14 dB, a factor of 25 in intensity, and it is the single biggest reason unamplified concerts are possible at all.

Levels of Real Musical Sounds

A rough table, for calibration. All at the listener's position in a concert hall:

Marking	Level (dB SPL)
pp	50–60
mf	70–80
ff	90–100

An orchestra's usable dynamic range is therefore about 40–50 dB, which is far less than the 120 dB the ear can handle. The limit at the bottom is the noise of the hall and the audience; at the top it is what the instruments can produce.

Note also that dynamic markings are **not** absolute levels. A *pianissimo* in a Mahler symphony and a *pianissimo* in a string quartet differ by a good deal, and performers read the marking as an instruction about relative level and character rather than about decibels.

7.4 Loudness Level: The Phon

Equal-Loudness Contours

Now the frequency dependence. The experiment is straightforward to describe: play a 1 kHz reference tone at a known level, then play a tone at some other frequency and ask the listener to adjust it until the two sound equally loud. Repeat across the spectrum, and across reference levels, and average over many listeners.

The result is the set of **equal-loudness contours**, standardized as ISO 226.

Three features of those curves carry most of the content.

The ear is most sensitive around 3–4 kHz. This is the ear-canal resonance of Chapter 6, and it is why that region is where speech intelligibility lives and where a singer must compete (Chapter 13).

Sensitivity falls sharply at low frequencies. At 50 Hz the threshold is about 45 dB SPL: a sound 45 dB above the 1 kHz threshold is only just audible.

The curves flatten as level rises. This is the important one, and it is the subject of the next subsection.

Reading the Contours

The **phon** is defined by those curves: a sound has a loudness level of N phons if it is judged as loud as a 1 kHz tone at N dB SPL.

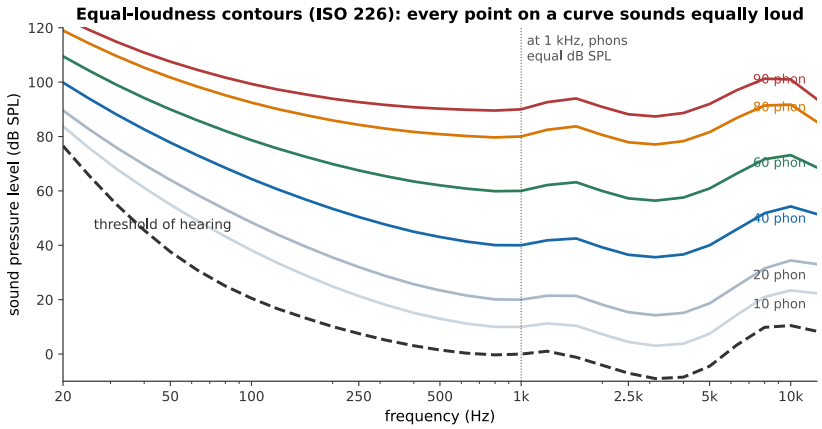


Figure 7.5: The equal-loudness contours. Every point on one curve sounds equally loud. The curves are labeled in **phons**: by definition, a tone has a loudness level of N phons if it sounds as loud as a 1 kHz tone at N dB SPL. At 1 kHz, therefore, phons and dB SPL agree exactly; everywhere else they do not.

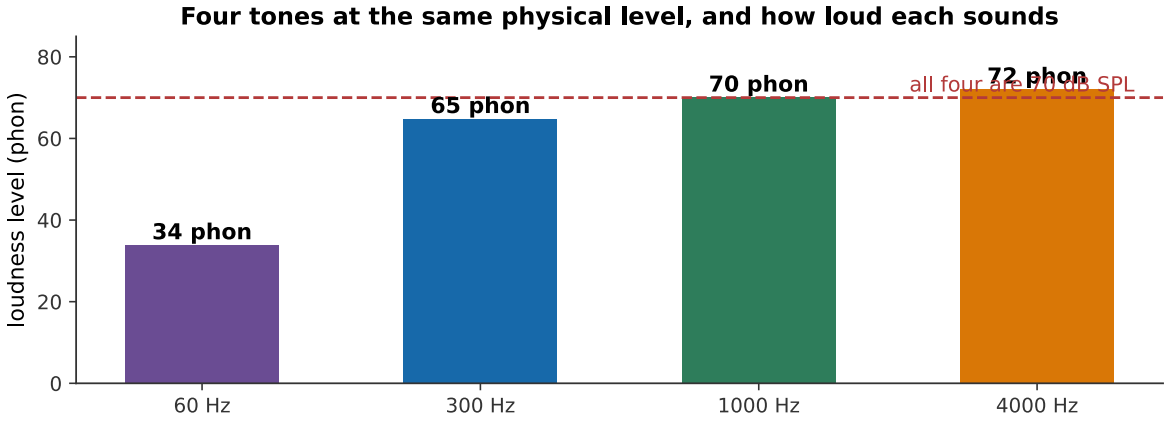


Figure 7.6: Four tones, all at the same sound pressure level, rising in frequency. They are emphatically not equally loud, the 60 Hz tone sounds around 34 phon while the 4 kHz tone sounds around 72 phon, a difference of nearly four doublings in loudness for identical physical sound. (Judge this at a moderate volume; the effect depends on level, which is exactly the point.) *What you hear:*Four tones in succession at identical physical level. The first, at 60 Hz, is faint; the second louder; the third and fourth clearly the loudest. They are the same level throughout. Listen: Equal Amplitude Tones

Why Quiet Music Loses Its Bass

Because the contours are not parallel, **the ear's frequency response depends on level**.

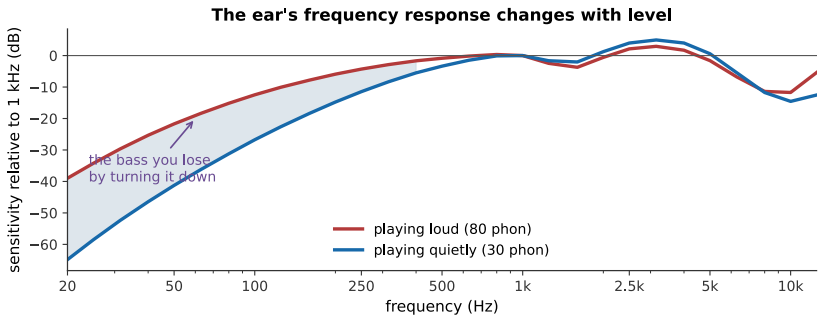


Figure 7.7: The same music, played loudly and quietly, relative to 1 kHz. Turning the volume down does not scale everything equally: the low frequencies fall away faster than the middle, because the contours crowd together at low levels.

The consequence is familiar to everyone. **Music played quietly sounds thin.** Turning the volume down does not merely make everything smaller; it changes the balance, taking the bass first.

> **Note**
> This is what the "loudness" button on old hi-fi amplifiers did. It applied a bass boost at low volume settings, attempting to compensate for the crowding of the contours. The idea is sound in principle and hard in practice, because the correct compensation depends on the *absolute* level at the listener's ears, which the amplifier has no way of knowing.

7.5 Loudness Itself: The Sone

Doubling Loudness

The phon says when two sounds are equally loud. It does not say when one is *twice* as loud, and it cannot: phons are defined by a matching procedure, and matching says nothing about ratios.

To get ratios, ask listeners directly: adjust this sound until it is twice as loud as that one. The answer, robustly reproduced, is

$$\text{an increase of 10 phons doubles the loudness.} \tag{7.10}$$

The **sone** encodes this. One sone is defined as 40 phons, and

$$S = 2^{(P-40)/10}, \tag{7.11}$$

with S in sones and P in phons. So 40 phons is 1 sone, 50 phons is 2 sones, 60 phons is 4 sones.

Ten phons is a factor of ten in intensity. So **a tenfold increase in intensity is only a doubling in loudness**, which is the compression that makes the ear's enormous range usable.

Worked example: how much louder is the orchestra?
A passage marked pp *is 55 dB SPL and one marked* ff *is 95 dB SPL, both broadband so that phons and dB SPL are roughly equal. How much louder does the* ff *sound?*

The difference is 40 phons, which is four doublings:

$$\frac{S_{ff}}{S_{pp}} = 2^{40/10} = 2^4 = 16. \tag{7.12}$$

Therefore, the *fortissimo* sounds about sixteen times as loud, from a sound that is 10^4, ten thousand times, more intense.

That compression is worth pausing on. It is what lets one pair of ears handle a whisper and an orchestra, and it is also why adding musicians is such an inefficient way to get louder.

Masking

A sound that would be perfectly audible on its own can be rendered completely inaudible by another sound. This is **masking**, and Chapter 6 supplied the mechanism: a masker's traveling wave excites a region of the basilar membrane, and anything whose own region falls inside that excitation is lost in it.

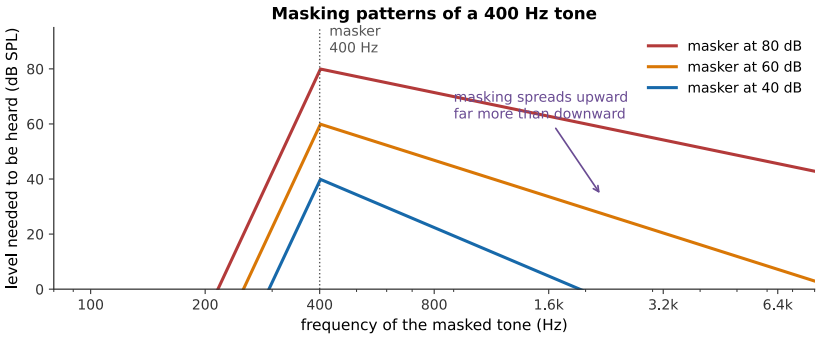


Figure 7.8: Masking patterns produced by a 400 Hz tone. Each curve shows the level another tone needs in order to be heard. Two features matter: masking is **asymmetric**, spreading far further upward in frequency than downward, and the asymmetry gets worse as the masker gets louder.

The upward spread of masking has direct musical consequences. A loud bass line buries a quiet high part far more than the reverse, so arrangers give the melody to a high instrument and a mix engineer who wants to hear the hi-hat turns the bass *down* rather than the hi-hat up.

It is also, remarkably, what makes MP3 possible. If a component of a signal is masked, a listener cannot hear it, so a codec can simply discard it and spend the saved bits elsewhere. Chapter 15 develops this.

The Critical Band

Masking is strongest when masker and masked are close in frequency, and the width of the region over which they interact strongly is the **critical band**.

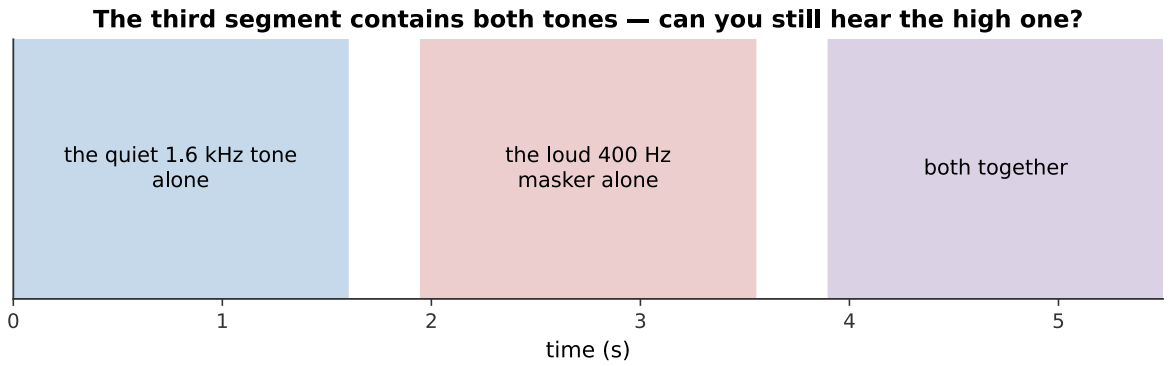


Figure 7.9: Masking in three segments. The high tone in the third segment is present at exactly the level it had in the first. Nothing has been removed from the signal; it has been removed from your perception of it. *What you hear:*A quiet high tone, clearly audible on its own. Then a loud low tone. Then both together, in which the high tone has become difficult or impossible to hear. Listen: Masking Demo

It is roughly 100 Hz wide below 500 Hz, and roughly a third of an octave above that, which means it is approximately **constant in musical terms**, a little wider than a major third, across most of the range. That is not a coincidence: the critical band corresponds to a roughly constant distance along the basilar membrane, and Chapter 6 showed that the membrane's map is logarithmic.

The critical band is the ear's frequency-resolution limit, and it does a great deal of work in what follows:

- Two tones within a critical band interfere audibly and sound **rough**; two tones further apart do not. This is the physical basis of dissonance (Chapter 8).
- Noise masks a tone only if the noise lies within about a critical band of it.
- Loudness adds differently within and across critical bands. Spread the same energy across several bands and it sounds louder than if it is concentrated in one; so a chord is louder than a single note of the same total intensity, and a full orchestration carries further than a doubled unison.

7.6 Summary

- **Intensity goes as pressure squared.** The decibel therefore has two forms: $10\log_{10}$ for intensity and power, $20\log_{10}$ for pressure and amplitude. They describe the same sound.
- **Hearing spans 10^{12} in intensity**, from 10^{-12} W/m^2 to about 1 W/m^2, referred to $p_0 = 20$ μPa. A decibel figure without a stated reference is a ratio, not a level.
- **Two facts do most decibel arithmetic**: a factor of ten is 10 dB, a factor of two is 3 dB (in intensity).
- **Incoherent sources add in intensity**, so n of them give $10\log_{10}(n)$ dB. Sixty violins are 18 dB above one, less than four times as loud.
- **Distance costs 6 dB per doubling** outdoors, and much less than that indoors beyond the critical distance.
- **Equal-loudness contours** describe the ear's frequency dependence. The **phon** is defined by matching against a 1 kHz tone, so phons and dB SPL agree at 1 kHz and nowhere

else. Sensitivity peaks near 3–4 kHz and collapses at low frequencies.
- **The contours flatten as level rises**, so the ear's frequency balance depends on volume. Quiet music loses its bass first.
- **The sone measures loudness itself**: 1 sone is 40 phons, and each 10 phons doubles it. A tenfold increase in intensity is a twofold increase in loudness.
- **Masking** renders an audible sound inaudible, spreading much further upward in frequency than downward. It is the basis of perceptual audio coding.
- **The critical band** is the ear's resolution limit, roughly a third of an octave, hence roughly constant in musical terms. It underlies roughness, dissonance, and the way loudness sums across a spectrum.

7.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple choice.** Eight identical, mutually incoherent sources play together. Approximately how much higher is their combined level than one source: 3, 8, 9, or 18 dB?

2. **True or false.** Doubling sound-pressure amplitude increases sound-pressure level by about 6 dB.

3. **Drag the words.** Complete the rules connecting one sone to phons, a loudness doubling to phons, an outdoor distance doubling to dB, and a doubling of incoherent sources to dB.

4. **Fill in the blanks.** Use _____ $\log_{10}$ for pressure or amplitude ratios, but _____ $\log_{10}$ for intensity or power ratios.

5. **Mark the words.** Identify the two ideas explaining why quiet playback loses bass: "At low levels the ear is less sensitive to low frequencies, and equal-loudness contours flatten as the overall level rises."

7.8 Conceptual Questions

1. Explain why the decibel scale uses $10 \log$ for intensity and $20 \log$ for pressure, and why both give the same level for the same sound.

2. A student writes "the sound was 70 dB". Explain what is missing from this statement and why it matters.

3. Explain why two loudspeakers fed from the same signal give +6 dB while two violinists playing the same note give +3 dB.

4. An orchestra has sixty violins and three trumpets. Use the addition rule to comment on whether this is a sensible balance.

5. Explain why music played quietly sounds thin, using the shape and spacing of the equal-loudness contours.

6. A 50 Hz tone and a 3 kHz tone are played at the same dB SPL. Which sounds louder, and roughly by how many phons at moderate levels?

7. Explain why a loud bass line masks a quiet cymbal much more than a quiet cymbal masks a loud bass line.

8. Explain how masking allows an audio codec to discard part of a signal without an audible change, and state what would go wrong if the codec's model of masking were inaccurate.

7.9 Problems

Exercise 7.1
(Straightforward) A sound has intensity 3.5×10^{-6} W/m^2. (a) What is its sound intensity level in dB? (b) What pressure amplitude does it correspond to, taking $\rho v = 413$ kg m^{-2} s^{-1}?

Solution to Exercise 7.1
(a) Using $L = 10 \log_{10}(I/I_0)$ with $I_0 = 10^{-12}$ W/m^2:

$$L = 10 \log_{10}\left(\frac{3.5 \times 10^{-6}}{1 \times 10^{-12}}\right) = 10 \log_{10}(3.5 \times 10^{6}) = 65.4 \text{ dB}. \tag{7.13}$$

(b) From $I = p^2/\rho v$:

$$p = \sqrt{I \rho v} = \sqrt{(3.5 \times 10^{-6})(413)} = \sqrt{1.446 \times 10^{-3}} = 0.038 \text{ Pa}. \tag{7.14}$$

Therefore, the level is 65 dB and the pressure amplitude about 38 mPa, roughly conversational.

Exercise 7.2
(Moderate) Without a calculator, find the level difference corresponding to an intensity ratio of (a) 100, (b) 8, (c) 40, (d) 1/2.

Solution to Exercise 7.2
Using $\times 10 \to 10$ dB and $\times 2 \to 3$ dB.
(a) $100 = 10 \times 10 \to 10 + 10 = 20$ dB.
(b) $8 = 2 \times 2 \times 2 \to 3 + 3 + 3 = 9$ dB.
(c) $40 = 10 \times 2 \times 2 \to 10 + 3 + 3 = 16$ dB.
(d) $1/2 \to -3$ dB.
Therefore: 20, 9, 16, and -3 dB. (Exact values: 20.0, 9.03, 16.02, -3.01.)

Exercise 7.3
(Moderate) One clarinet produces 72 dB at a listener. (a) What level do four identical clarinets produce? (b) How many clarinets would be needed for 82 dB? (c) Comment on the practicality.

Solution to Exercise 7.3

(a) Four incoherent sources add $10\log_{10}(4) = 6.0$ dB:

$$L = 72 + 6 = 78 \text{ dB}. \tag{7.15}$$

(b) A rise of 10 dB needs

$$10\log_{10}(n) = 10 \;\Rightarrow\; n = 10. \tag{7.16}$$

(c) Ten clarinets for 10 dB, which is one doubling of loudness. Getting a second doubling would take a hundred. An ensemble cannot get much louder by recruiting, so composers reach for instruments with intrinsically higher output, brass, rather than for more of the same.

Exercise 7.4

(Straightforward) A singer produces 88 dB at 1 m. Outdoors, find the level at (a) 2 m, (b) 8 m, (c) 50 m.

Solution to Exercise 7.4

Using $L_2 = L_1 - 20\log_{10}(r_2/r_1)$.

(a) $88 - 20\log_{10}(2) = 88 - 6.0 = 82$ dB.

(b) $88 - 20\log_{10}(8) = 88 - 18.1 = 70$ dB.

(c) $88 - 20\log_{10}(50) = 88 - 34.0 = 54$ dB.

Therefore, 82, 70, and 54 dB. At 50 m outdoors the singer is at about the level of a quiet conversation, outdoor performance essentially requires amplification; indoor performance does not.

Exercise 7.5

(Straightforward) A tone at 100 Hz must be played at 50 dB SPL to sound as loud as a 1 kHz tone at 30 dB SPL. (a) What is the loudness level of each, in phons? (b) How much more intense is the 100 Hz tone?

Solution to Exercise 7.5

(a) The loudness level is defined by the 1 kHz match, so **both are 30 phons**.

(b) The level difference is $50 - 30 = 20$ dB, so the intensity ratio is

$$10^{20/10} = 100. \tag{7.17}$$

Therefore, the 100 Hz tone must be a hundred times more intense to sound equally loud. This is exactly the message of the equal-loudness contours: physical equality and perceptual equality are different things.

Exercise 7.6

(Moderate) A sound has a loudness level of 70 phons. (a) What is its loudness in sones? (b) What loudness level would sound twice as loud? (c) Four times?

Solution to Exercise 7.6

(a) Using $S = 2^{(P-40)/10}$:

$$S = 2^{(70-40)/10} = 2^3 = 8 \text{ sones}. \tag{7.18}$$

(b) Twice as loud is 16 sones, which is 2^4, so $P = 40 + 40 = 80$ phons.

(c) Four times is 32 sones, 2^5, so $P = 90$ phons.

Therefore, 8 sones; doubling needs +10 phons and quadrupling +20. Note that quadrupling the loudness requires a hundredfold increase in intensity.

Exercise 7.7

(Moderate) A rock concert measures 112 dB. (a) How many times more intense is this than the 85 dB eight-hour limit? (b) Using the 3 dB exchange rate, what is the safe exposure time?

Solution to Exercise 7.7

(a) The difference is $112 - 85 = 27$ dB:

$$10^{27/10} = 501. \tag{7.19}$$

(b) 27 dB is nine steps of 3 dB, so nine halvings of eight hours:

$$T = \frac{8 \text{ h}}{2^9} = \frac{8}{512} \text{ h} = 0.0156 \text{ h} = 56 \text{ seconds}. \tag{7.20}$$

Therefore, the sound is about 500 times more intense than the limit, and the safe unprotected exposure is under a minute. A three-hour concert at that level is roughly two hundred times the permitted daily dose.

Exercise 7.8

(Moderate) Two sounds of 70 dB and 76 dB are played together. (a) What is the combined level? (b) What is it if the second is 60 dB instead? (c) State the general rule your second answer illustrates.

Solution to Exercise 7.8

Intensities add, so convert, add, and convert back.

(a) Relative to I_0: $10^{7.0} = 1.00 \times 10^7$ and $10^{7.6} = 3.98 \times 10^7$. The sum is 4.98×10^7, so

$$L = 10 \log_{10}(4.98 \times 10^7) = 77.0 \text{ dB}. \tag{7.21}$$

(b) $10^{7.0} = 1.00 \times 10^7$ and $10^{6.0} = 0.10 \times 10^7$. The sum is 1.10×10^7:

$$L = 10 \log_{10}(1.10 \times 10^7) = 70.4 \text{ dB}. \tag{7.22}$$

(c) **A sound 10 dB below another adds almost nothing**, less than half a decibel. Anything more than about 10 dB down can usually be ignored when adding levels, which is a useful shortcut and also the reason a single loud instrument can dominate an ensemble entirely.

Exercise 7.9

(Moderate) An orchestra plays a passage at 95 dB SPL and a passage at 60 dB SPL. Treating both as broadband so phons and dB SPL agree: (a) Find the loudness of each in sones. (b) Find the ratio. (c) Find the intensity ratio, and comment.

Solution to Exercise 7.9

(a) $S_{95} = 2^{(95-40)/10} = 2^{5.5} = 45.3$ sones; $S_{60} = 2^{(60-40)/10} = 2^2 = 4$ sones.

(b) The loudness ratio is $45.3/4 = 11.3$.

(c) The intensity ratio is $10^{35/10} = 3.2 \times 10^3$.

Therefore, a sound over three thousand times more intense sounds about eleven times louder. The compression is enormous, roughly a cube root, and it is what allows a single dynamic range to accommodate both a solo flute and a full orchestra.

Exercise 7.10

(Moderate) The critical band is about a third of an octave above 500 Hz. (a) How wide is the critical band at 1 kHz, in hertz? (b) At 4 kHz? (c) Express each in semitones.

Solution to Exercise 7.10

A third of an octave is a frequency ratio of $2^{1/3} = 1.26$.

(a) A band centered on 1 kHz running from $f/\sqrt[6]{2}$ to $f\sqrt[6]{2}$:

$$1000/1.122 = 891 \text{ Hz} \quad \text{to} \quad 1000 \times 1.122 = 1122 \text{ Hz}, \tag{7.23}$$

a width of 231 Hz.

(b) The same fractional width at 4 kHz gives 3565 to 4487 Hz, a width of 922 Hz.

(c) A third of an octave is four semitones in both cases, since the band is defined as a fixed ratio.

Therefore, the critical band is four times wider in hertz at 4 kHz than at 1 kHz, but identical in musical terms: a major third. That constancy in musical units is what makes the critical band useful for explaining consonance (Chapter 8).

Exercise 7.11

(Challenging) Show that if n incoherent sources each of intensity I combine, the level rise is $10\log_{10}(n)$, and explain why the corresponding result for n coherent sources is $20\log_{10}(n)$.

Solution to Exercise 7.11

Incoherent. The phases are uncorrelated, so the cross terms in the squared sum average to zero over time, and the intensities add: $I_{\text{tot}} = nI$. The level rise is

$$10\log_{10}\left(\frac{nI}{I}\right) = 10\log_{10}(n). \tag{7.24}$$

Coherent. The phases are locked, so the *pressures* add: $p_{\text{tot}} = np$. Since intensity goes as pressure squared, $I_{\text{tot}} = n^2 I$, and

$$10\log_{10}(n^2) = 20\log_{10}(n). \tag{7.25}$$

Therefore, coherent addition gives exactly twice the decibel rise of incoherent addition. For $n = 2$ that is 6 dB against 3 dB: the difference between two loudspeakers and two violinists.

The coherent result cannot hold everywhere in space: constructive interference at one point must be paid for by destructive interference elsewhere (Chapter 3). The 6 dB applies on the axis where the two arrive in step.

Exercise 7.12
(Moderate) A 60 Hz tone and a 3 kHz tone are both played at 70 dB SPL. Using the equal-loudness contours, estimate the loudness level of each in phons, and then the loudness of each in sones. What is the ratio?

Solution to Exercise 7.12
Reading the contours of Figure 7.5 at 70 dB SPL:

At 60 Hz the 70 dB point lies near the 35 phon contour. At 3 kHz it lies a little above the 70 phon contour, call it 75 phon, since the ear is *more* sensitive than at 1 kHz there.

In sones:

$$S_{60} = 2^{(35-40)/10} = 2^{-0.5} = 0.71, \qquad S_{3000} = 2^{(75-40)/10} = 2^{3.5} = 11.3. \tag{7.26}$$

The ratio is $11.3/0.71 = 16$.

Therefore, two sounds of *identical physical level* differ by a factor of about sixteen in loudness, four doublings. Any claim about how loud something "is", made without reference to its frequency content, is close to meaningless.

Exercise 7.13
(Moderate) A mixing engineer boosts a vocal by 3 dB. (a) By what factor has its intensity increased? (b) By what factor its pressure amplitude? (c) By roughly what factor its loudness? (d) Comment on which of these the engineer is thinking about.

Solution to Exercise 7.13
(a) Intensity: $10^{3/10} = 2.0$.

(b) Pressure: $10^{3/20} = 1.41$.

(c) Loudness: 3 dB is 0.3 of a 10-phon doubling, so

$$2^{0.3} = 1.23, \tag{7.27}$$

about a 23% increase.

(d) None of them explicitly, but (c) is the one that matters. A 3 dB boost doubles the energy and is barely a noticeable change in loudness, so mixing moves are typically made in steps of 1 to 3 dB and the smallest audible change in a mix context is often quoted as about 1 dB.

Exercise 7.14
(Challenging) A hall has a background noise level of 30 dB SPL. A solo flute produces 58 dB at the back row. (a) What is the combined level? (b) What is the signal-to-noise ratio? (c) The audience arrives and the background rises to 42 dB. Recompute both, and comment on what the audience has cost the flutist.

Solution to Exercise 7.14
(a) Relative to I_0: flute $10^{5.8} = 6.31 \times 10^5$, noise $10^{3.0} = 1.00 \times 10^3$. Sum 6.32×10^5:

$$L = 10 \log_{10}(6.32 \times 10^5) = 58.0 \text{ dB}. \tag{7.28}$$

(b) The signal-to-noise ratio is $58 - 30 = 28$ dB.

(c) With noise at 42 dB: $10^{4.2} = 1.58 \times 10^4$. Sum with the flute: $6.31 \times 10^5 + 0.158 \times 10^5 = 6.47 \times 10^5$, so $L = 58.1$ dB, essentially unchanged.

The signal-to-noise ratio, however, falls to $58 - 42 = 16$ dB.

Therefore, the audience has not made the hall measurably louder but has removed 12 dB of the flutist's headroom. The quietest passages the player could previously shape are now within a few decibels of the noise floor and cease to be audible as music. Halls are built with such low background-noise specifications for that reason, and a *pianissimo* is the first thing a noisy audience destroys.

Chapter 8

Pitch, Beats, Consonance, and Dissonance

Learning Objectives

By the end of this chapter, you should be able to:

- Distinguish pitch from frequency, and give two situations in which the pitch a listener reports is not the frequency present.
- Describe how pitch discrimination varies across the audible range, and state the approximate just-noticeable difference in frequency.
- Derive the beat frequency for two tones of nearby frequency, and explain how a piano tuner uses beats to set an interval.
- Explain the missing fundamental, and use it to argue that pitch cannot be read off the place of maximum excitation alone.
- Compare place and temporal theories of pitch, and state what each explains and where each fails.
- Explain consonance and dissonance in terms of roughness between partials falling within a critical band.
- Predict, from the spectra of two complex tones, which intervals between them will sound rough and which smooth, and relate the answer to small whole-number frequency ratios.
- Explain why the consonance of an interval depends on the timbre of the instrument playing it, and give an example.
- Identify combination tones and describe the conditions under which they are audible.

Introduction

This chapter asks two questions that look simple and are not.

What determines the pitch of a sound? The obvious answer, its frequency, is wrong often enough to be interesting. A sound can have a clear pitch at a frequency that is entirely absent from it. Two sounds of the same frequency can have different pitches, depending on how loud they are. And above about 5 kHz, pitch perception degrades badly even though hearing continues for another two octaves.

Why do some pairs of notes sound good together? This is the older question. Pythagoras is supposed to have noticed, two and a half millennia ago, that strings whose lengths are in simple whole-number ratios sound well together, and that observation has been the foundation of Western music theory ever since. But *why* should the ratio $3:2$ be pleasant

and the ratio 45 : 32 not? The numbers themselves cannot be doing anything. Something about the ear must be.

The two questions turn out to have a common answer, and it is the one Chapter 7 supplied: the **critical band**. Both pitch and consonance are consequences of the ear resolving frequencies with a limited but definite sharpness, and of what happens when two components fall inside one resolution element.

This chapter is also where the book stops being purely deductive. The results here are experimental facts about listeners, and the theories that explain them are good but incomplete. Where a theory fails, the text says so.

8.1 Pitch and Frequency

The Pitch of a Pure Tone

For a pure tone in the middle of the audible range, pitch tracks frequency closely and monotonically. Double the frequency and listeners agree the pitch has risen by an octave. That much is uncontroversial.

But pitch depends slightly on **level** as well. A pure tone below about 1 kHz is heard as slightly flatter when played louder; one above about 2 kHz slightly sharper. The effect is small, a few percent at extreme levels, and it does not arise for complex tones, so it never troubles musicians. It remains decisive evidence that pitch is computed rather than measured.

Just-Noticeable Differences

How finely can the ear tell two frequencies apart? The answer depends on where you ask.

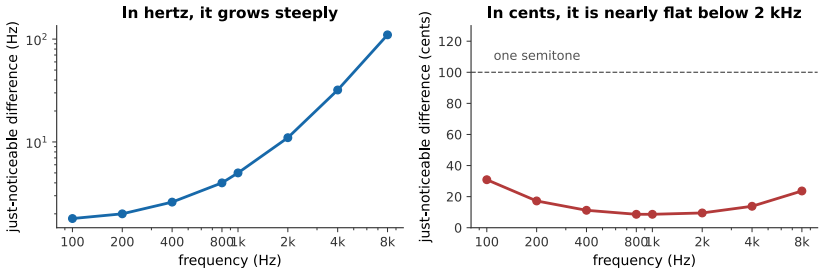


Figure 8.1: The just-noticeable difference in frequency. Expressed in hertz it grows steeply with frequency; expressed in **cents**, hundredths of a semitone, it is nearly constant at about 3–6 cents across the whole musically useful range, and only deteriorates above about 2 kHz.

That flatness in cents is the important result, and it reflects the logarithmic layout of the cochlea from Chapter 6. **The ear's pitch resolution is proportional, not absolute**, which is exactly why music is built on ratios.

Three to six cents is extraordinarily fine. A semitone is 100 cents, so the ear resolves about one thirtieth of a semitone, far finer than any performer needs, and far finer than the differences between tuning systems that Chapter 9 will agonize over. When that chapter says a tuning system is 14 cents out, it means an error two to four times the threshold of detection.

> **Note**
> Pitch discrimination is much finer for *successive* tones than the ability to identify a pitch in isolation. Most people can hear that two notes played one after another differ by five cents, while being quite unable to say what either note was. Absolute pitch, naming a note with no reference, is a different and much rarer ability, and it is not what these measurements test.

Where Pitch and Frequency Part Company

Three situations break the simple correspondence, and each is instructive.

Above about 5 kHz, pitch perception deteriorates sharply. Listeners cannot reliably recognize melodies played in pure tones above this, and cannot tune intervals there. Chapter 6 gave the mechanism: phase locking in the auditory nerve fails above 4–5 kHz, so the timing information disappears and only place is left. The top note of a piano is 4186 Hz, and that is very likely not a coincidence.

Very short tones have no definite pitch. A tone must last several cycles before a pitch can be assigned, and below about 10 ms it is heard as a click. This matters for percussion and for the very lowest notes, where ten cycles of a 30 Hz tone takes a third of a second.

A complex tone can have a pitch at a frequency it does not contain. This is the big one, and §8.3 is about it.

8.2 Beats

Two Nearby Frequencies

Play two tones of nearly equal frequency together. The result is a single tone, at roughly the average frequency, whose loudness rises and falls at a slow, steady rate.

The reason is superposition (Chapter 3) in time rather than in space. Two tones at 220 and 223 Hz start in step, but the second completes three more cycles per second, so after a sixth of a second it is half a cycle ahead and they cancel; after a third of a second it is a full cycle ahead and they reinforce again.

The Beat Equation

The trigonometry is one line:

$$\sin(2\pi f_1 t) + \sin(2\pi f_2 t) = 2\,\cos\left(2\pi \frac{f_2 - f_1}{2} t\right) \sin\left(2\pi \frac{f_1 + f_2}{2} t\right). \tag{8.1}$$

This is a tone at the **average** frequency, multiplied by a slowly varying envelope at **half the difference** frequency. The envelope goes through zero twice per cycle of that cosine, so the loudness peaks

$$f_{\text{beat}} = |f_2 - f_1| \tag{8.2}$$

times per second.

What is heard as the rate changes is not one phenomenon but a sequence of them:

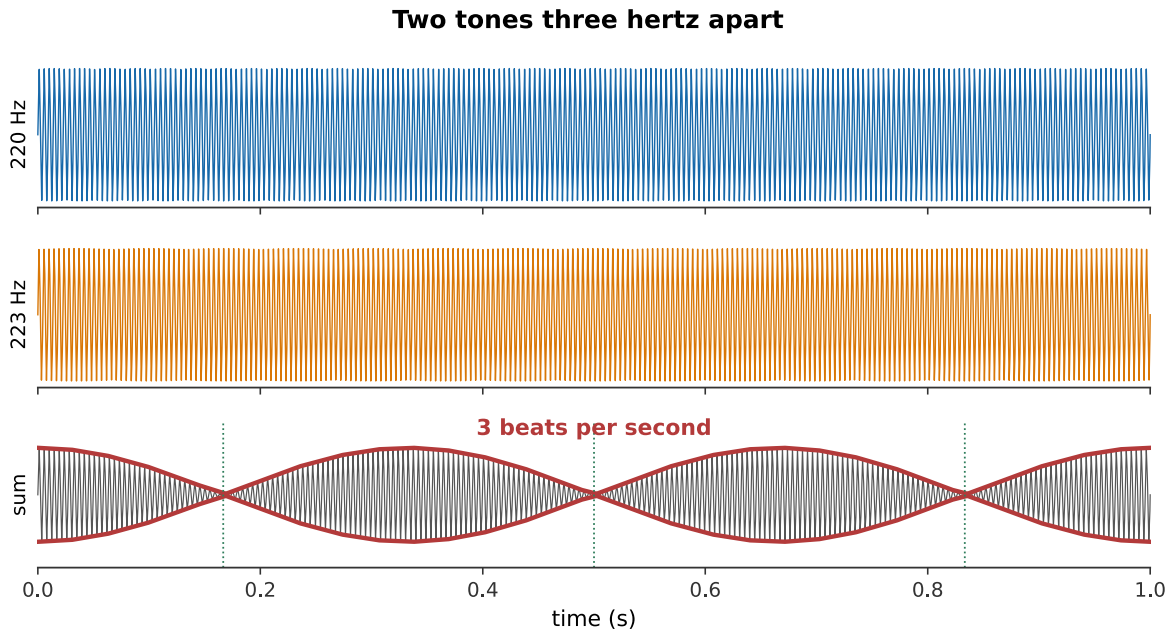


Figure 8.2: Two nearby tones, and their sum, slowed down here so the beat is visible — real musical beats like 220 and 223 Hz oscillate too fast to watch directly. The two drift in and out of step, reinforcing when they are aligned and canceling when they are opposed. In the original, the envelope completes three cycles per second, the difference between the two frequencies.

Difference	What is heard
0–6 Hz	A single tone, throbbing. Beats can be counted.
6–20 Hz	Too fast to count. A rough, fluttering quality.
20 Hz to a critical band	**Roughness**, harsh, unpleasant, unmistakable.
Beyond a critical band	Two separate smooth tones. The roughness vanishes.

Tuning by Beats

Beats are how instruments are actually tuned, and the method is far more precise than listening for "sameness".

When two strings are close, the ear cannot tell which is higher. But it can count beats, and the beat rate goes to zero at the exact match. A tuner tightens one string until the beats slow, slows further, and stop. Since the ear can detect a beat rate down to a fraction of a hertz, this places the unison within a small fraction of a cent, far better than the 3–6 cent discrimination of §8.1.

The method extends to other intervals, and this is what makes Chapter 9 practical. Tune a fifth: the third harmonic of the lower note and the second harmonic of the upper should coincide at a perfect $3:2$. If the fifth is slightly narrow, those two harmonics differ slightly and beat. **A piano tuner setting equal temperament is counting a specified number**

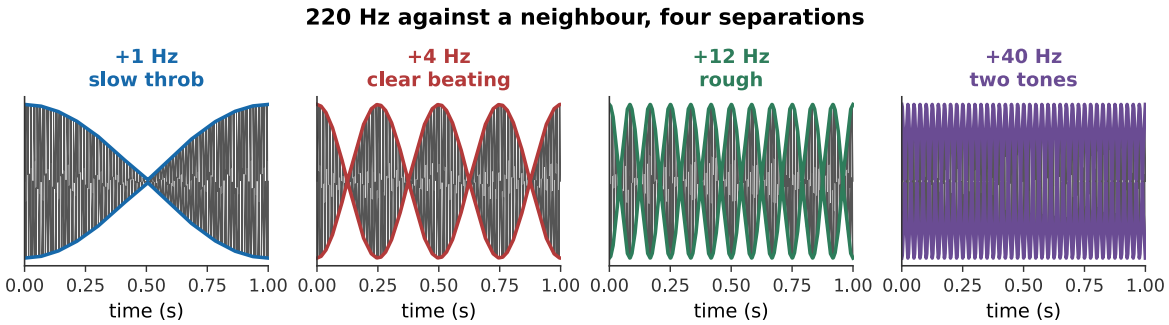


Figure 8.3: A 220 Hz tone against neighbors 1, 4, 12, and 40 Hz away. Listen for the transition between the third and fourth: the roughness does not fade gradually, it *stops* once the two tones are far enough apart to fall in different critical bands. That transition is the whole mechanism of §8.4. *What you hear:*Four pairs of tones in succession. The first throbs about once a second; the second beats clearly; the third is rough and buzzy; the fourth has separated into two distinct smooth tones. Listen: Beat Rates

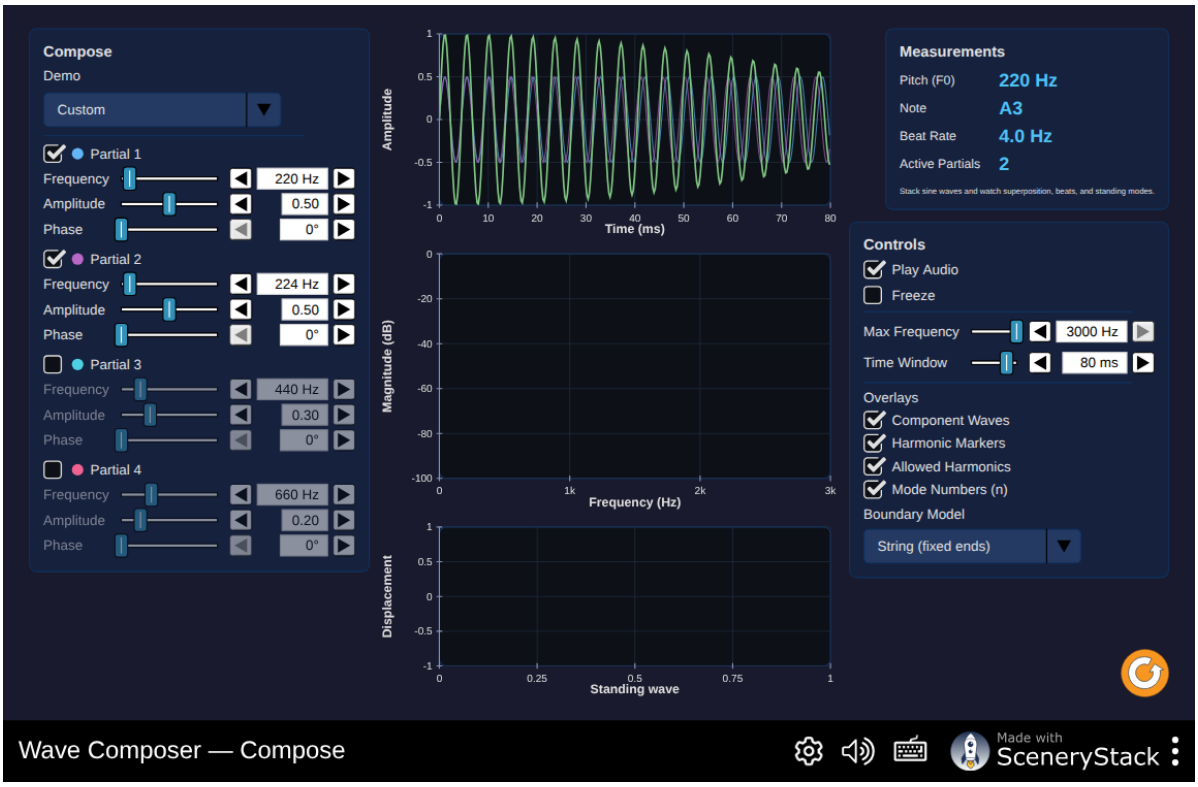


Figure 8.4: Add two sinusoids and vary the gap between them. Watch the envelope of the sum while counting the beats, and confirm that the rate is the *difference* of the two frequencies and not half of it, the factor of two of Exercise 8.9, visible rather than derived. Interactive simulation: Wave Composer

of beats per second on each interval, for a tempered fifth near middle C, about one beat every two seconds.

Worked example: tempering a fifth by beats

A tuner sets $A_3 = 220$ Hz and wants a tempered fifth above it. The equal-tempered fifth is 700 cents. How many beats per second should be heard?

The equal-tempered fifth is

$$f = (220\text{ Hz}) \times 2^{700/1200} = (220\text{ Hz})(1.49831) = 329.63\text{ Hz}. \tag{8.3}$$

A pure fifth would be $(220\text{ Hz}) \times 1.5 = 330.00$ Hz.

The beat is heard between the **third harmonic of the lower note** and the **second harmonic of the upper**:

$$3 \times (220\text{ Hz}) = 660.00\text{ Hz}, \qquad 2 \times (329.63\text{ Hz}) = 659.26\text{ Hz}. \tag{8.4}$$

$$f_{\text{beat}} = 660.00\text{ Hz} - 659.26\text{ Hz} = 0.74\text{ Hz}. \tag{8.5}$$

Therefore, the tuner should hear about three beats every four seconds. That is a countable rate, which is exactly why the method works, and it is why a piano tuner works up the scale listening rather than reading a meter.

8.3 The Pitch of a Complex Tone

The Missing Fundamental

Take a tone with harmonics at 200, 300, 400, 500, and 600 Hz. It has a pitch of 100 Hz, even though there is no 100 Hz component in it at all, and even if the sound is played through a system that cannot reproduce 100 Hz.

This is not a laboratory curiosity. It is happening constantly:

- A telephone passes nothing below 300 Hz, yet a male voice with a 110 Hz fundamental sounds perfectly normal on it.
- A small loudspeaker cannot reproduce 40 Hz, yet a bass guitar's low E is recognizable through it.
- Organ builders exploit it deliberately: a **resultant** stop sounds two pipes a fifth apart, and listeners hear a note an octave below the lower one, which no pipe in the building is producing.

Virtual Pitch and Pattern Matching

The explanation now generally accepted is that the auditory system **finds the fundamental that best explains the partials it is receiving**.

Given components at 200, 300, 400, 500 Hz, the system searches for a fundamental of which all are harmonics. 100 Hz works, they are its 2nd, 3rd, 4th, and 5th harmonics, and nothing higher does. So 100 Hz is reported.

This is a **pattern-matching** operation, and it is doing a genuinely useful job. In a room full of sound, the ear receives dozens of frequency components at once, from several sources.



Figure 8.5: Piano technician Dan Levitan demonstrates the audible beats in an imperfect unison and the clean result as its strings converge on the same frequency. The changing beat rate turns the equation above into a practical tuning method. Watch on YouTube: Unison Tuning

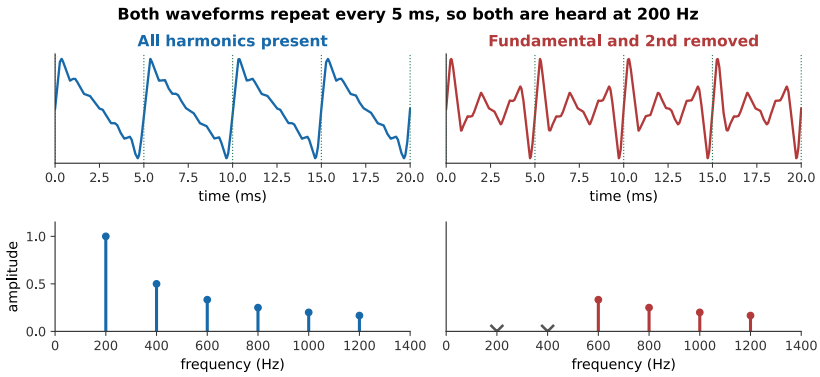


Figure 8.6: The missing fundamental. Removing the lowest partials changes the waveform's shape but not its **period**, and the pitch follows the period. The dotted lines mark 5 ms intervals; both waveforms repeat on them.

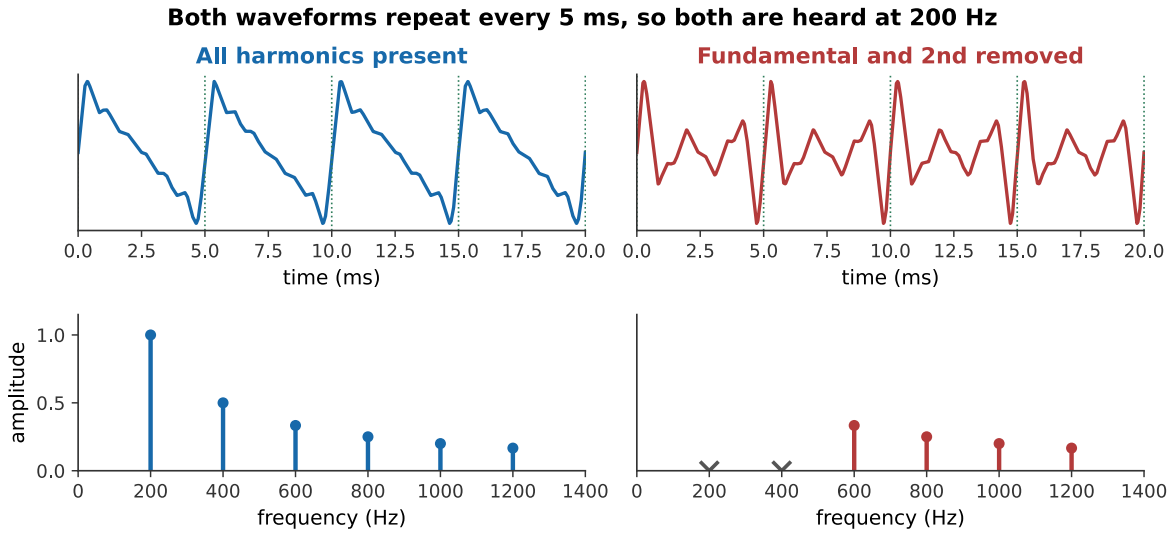


Figure 8.7: A 150 Hz tone, complete and with its two lowest partials removed. The *timbre* changes considerably; the *pitch* does not move at all. If pitch were simply the lowest frequency present, the second clip would sound an octave and a fifth higher. *What you hear:*Two tones. The second is thinner and less full-bodied than the first, but both are unmistakably the same note, the pitch does not jump up when the lowest partials are removed. Listen: All harmonics, Fundamental and 2nd removed

Grouping them into harmonic families is how it decides which components belong to which source, the classic "cocktail party" problem, and the perceived pitch is a by-product of that grouping.

Two pieces of evidence make it clear the mechanism is not merely arithmetic on the lowest partial:

Mistuning one harmonic shifts the pitch. Take harmonics at 200, 300, 400 Hz and move the 400 to 420. The perceived pitch rises slightly, to about 103 Hz, as though the system had compromised among competing candidates. A rule that merely computed the difference between adjacent partials would not do this.

The partials can be split between the ears. Present odd harmonics to one ear and even harmonics to the other and a single pitch is still heard at the fundamental. Whatever is doing the computation is in the brain, downstream of both cochleas.

Place Theory, Temporal Theory, and the Evidence

Two mechanisms have long competed to explain pitch, and the honest summary is that the auditory system uses both.

Place theory says pitch is read off *where* on the basilar membrane the excitation peaks. It explains why pitch perception survives at high frequencies, it explains masking, and it is undeniably part of the story because the tonotopic map exists.

Its problems are serious. It cannot easily account for the missing fundamental, since there is no excitation at the missing place. And it predicts far worse frequency resolution than listeners achieve: Chapter 6 put a critical band at about four semitones, while the ear resolves four *cents*.

Temporal theory says pitch is read off the *timing* of nerve firings, which phase-lock to the waveform. It explains the missing fundamental immediately, the waveform's period is

unchanged by removing the fundamental, and it explains the fine resolution, since timing can be measured very precisely.

Its problem is equally serious: phase locking fails above 4–5 kHz, and pitch perception, while degraded, does not vanish.

The evidence points to a division of labor. Below about 5 kHz, where music lives, timing dominates and gives fine resolution and the missing fundamental. Above it, only place is available, and performance degrades accordingly, which is exactly the pattern observed.

8.4 Consonance and Dissonance

Roughness and the Critical Band

Now Pythagoras's question. The modern answer, due in its quantitative form to Plomp and Levelt in 1965, is this:

> **Two partials sound rough when they are close enough to fall within one critical band, but not so close that they merely beat slowly.**

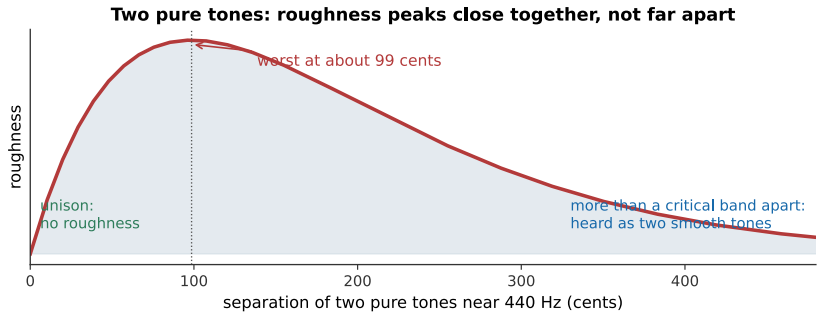


Figure 8.8: Roughness between two **pure** tones. At unison there is none; it rises to a maximum when the two are separated by roughly a quarter of a critical band, here about 70 cents, less than a semitone, and dies away once they are more than a critical band apart.

That figure is easily misread, so be explicit: **for two pure tones, the most dissonant interval is a small one**, around a semitone, and everything wider is progressively smoother. There is nothing special about a fifth, nothing special about an octave, and no preference for simple ratios anywhere in it.

Which raises the obvious objection. Music is not made of pure tones.

Why Simple Ratios Sound Smooth

Real musical tones have many partials, and the roughness of two complex tones is the **sum of the roughnesses of every pair of partials**.

Now the simple ratios earn their reputation. Two tones a perfect fifth apart, with harmonic partials, have partials at

$$\text{lower: } f,\ 2f,\ 3f,\ 4f,\ 5f,\ 6f, \ldots \tag{8.6}$$

$$\text{upper: } 1.5f,\ 3f,\ 4.5f,\ 6f, \ldots \tag{8.7}$$

The upper tone's 2nd partial coincides *exactly* with the lower tone's 3rd. Its 4th coincides with the lower's 6th. Coinciding partials contribute no roughness at all, and every coincidence is one fewer clashing pair.

Detune the fifth slightly and those coincidences become near-misses, pairs of partials a few hertz apart, which is exactly the recipe for beating and roughness.

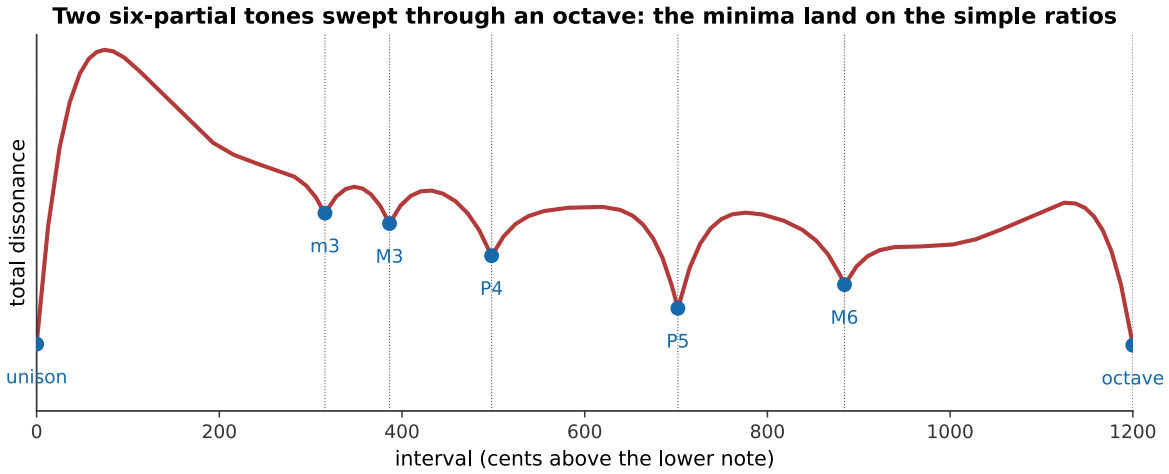


Figure 8.9: Total dissonance between two six-partial harmonic tones, computed by summing the roughness of every pair of partials as the upper note sweeps up an octave. **The minima fall on the simple ratios**, and nobody put them there. The curve was computed from a model of the ear's critical bands, with no musical knowledge in it at all.

That figure is, to my mind, the most satisfying result in the book. The consonant intervals of Western music emerge from a model containing nothing but the ear's frequency resolution and the fact that instruments produce harmonic partials.

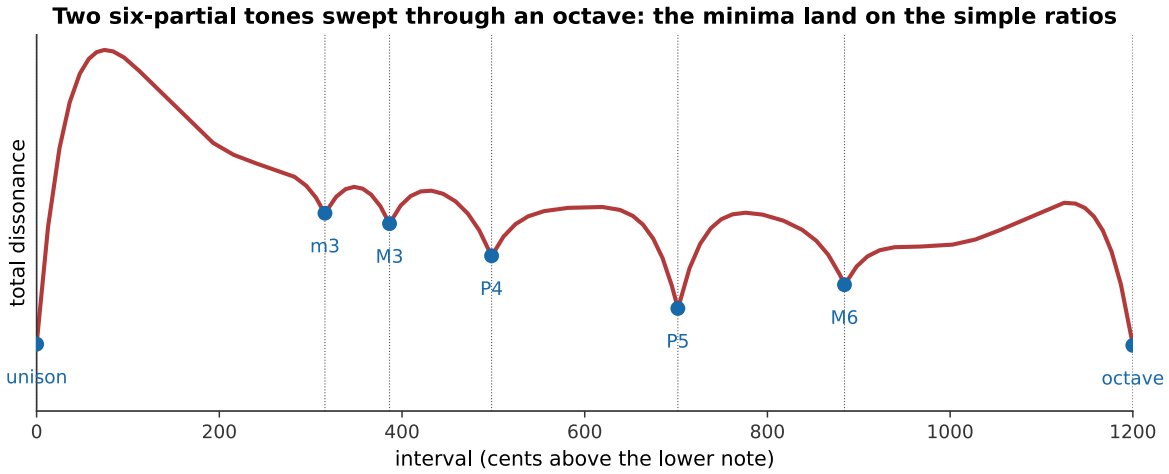


Figure 8.10: The same sweep, heard rather than computed. Follow it against the figure. The smooth points are not marked by anything in the sound generator, it simply glides, and yet they arrive on the fourth, the fifth, and the octave. *What you hear:*Two tones, the upper gliding slowly up an octave over sixteen seconds. The sound roughens and smooths repeatedly; the smooth points arrive at the familiar consonant intervals, and the roughest region is just above the unison. Listen: Interval Sweep

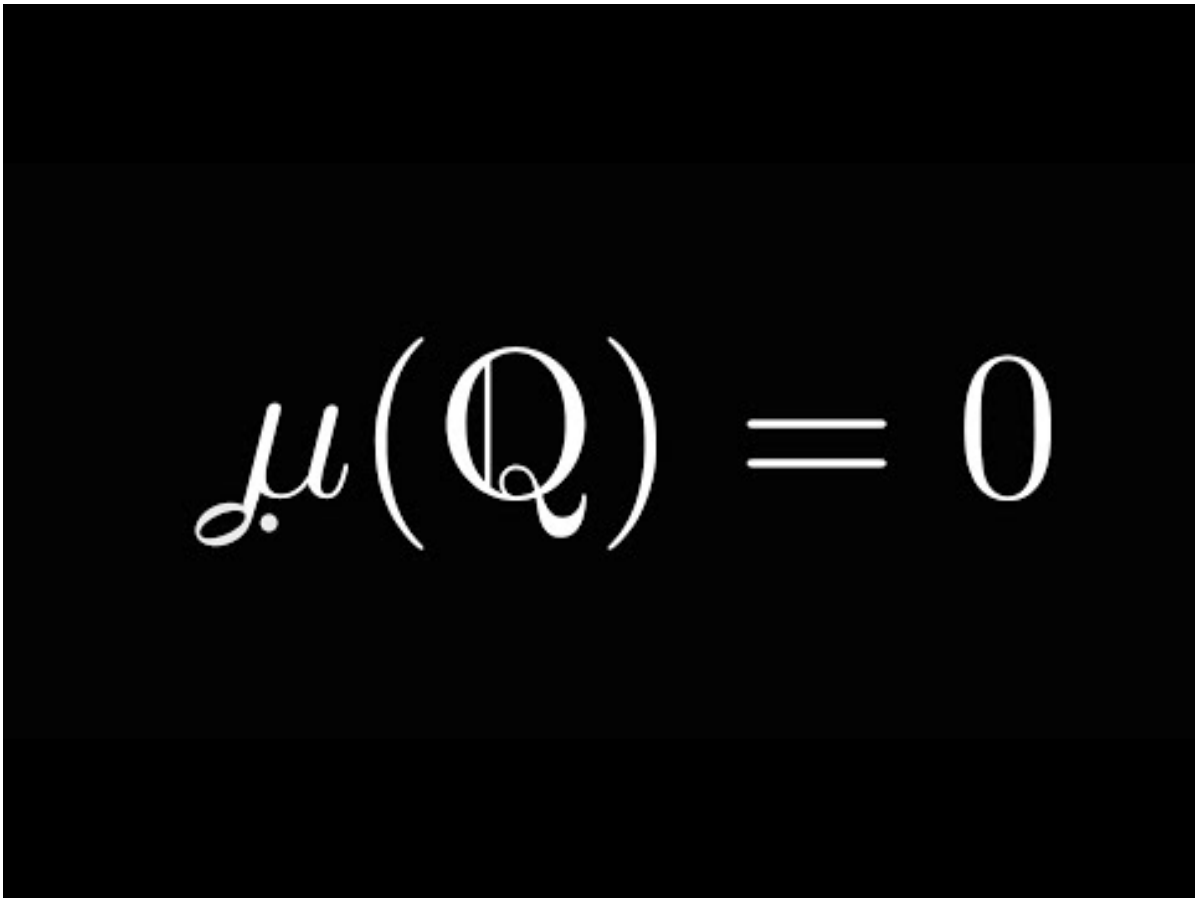


Figure 8.11: 3Blue1Brown approaches consonance through rational numbers and measure theory. It complements the chapter's physical account of beating and roughness by showing why simple frequency ratios occupy a mathematically special place. Watch on YouTube: Music and Measure Theory

Dissonance Curves

The curve above is specific to **six harmonic partials falling as** 0.88^n. Change the timbre and the curve changes.

- **More partials** deepen and sharpen the minima, because there are more coincidences to be had. Rich timbres make consonance more definite.
- **Fewer partials** flatten the curve. Two pure tones give no minima at all, as the roughness figure showed.
- **Inharmonic partials** move the minima somewhere else entirely.

Consonance Depends on the Instrument

That last point deserves its own demonstration, because it overturns something most musicians assume.

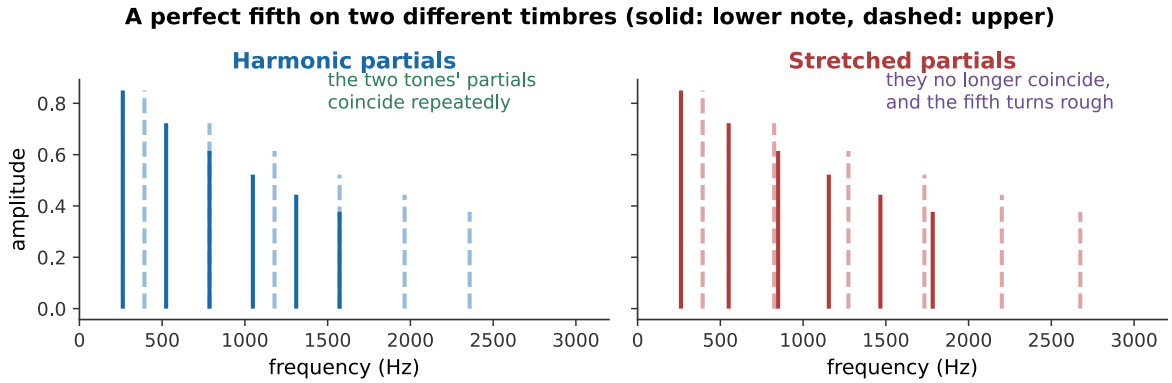


Figure 8.12: A perfect fifth, an exact $3:2$, played on harmonic partials and on stretched ones. The *ratio* is identical in both. What changes is whether the partials of the two tones coincide, and with the stretched timbre they no longer do. *What you hear:*The same interval twice. On the first timbre it is clean and settled; on the second, whose partials have been stretched, the same 3:2 ratio sounds distinctly rough and unresolved. Listen: Harmonic fifth, Stretched-partial fifth

The conclusion, which is William Sethares's, is worth stating plainly: **consonance is not a property of a frequency ratio. It is a relationship between a ratio and a timbre.** Simple ratios sound consonant on instruments with harmonic partials, and instruments with harmonic partials are what strings and air columns produce (Chapter 3). The Western scale and the Western orchestra fit each other because they grew up together.

This also explains, without any appeal to cultural relativism or its opposite, why the gamelan sounds out of tune to a Western ear and perfectly in tune to a Javanese one. Gamelan instruments are metallophones and gongs, whose partials are **inharmonic** (Chapter 12). Their dissonance curve has minima in different places, and the slendro and pelog scales sit on those minima. Both traditions are doing the same thing, choosing intervals that minimize roughness for their instruments, and arriving at different answers because their instruments are different.

8.5 Combination Tones and the Limits of the Theory

Combination Tones

Play two loud tones and a listener may hear additional tones that are not present in the signal at all. The most prominent is the **difference tone** at $f_2 - f_1$; others appear at $2f_1 - f_2$ and elsewhere.

These are not beats. Beats are a variation in the loudness of a tone at the average frequency; combination tones are heard as separate tones at genuinely different pitches.

The cause is **nonlinearity** in the ear. The cochlear amplifier of Chapter 6 is not a linear device, and any nonlinearity applied to a sum of two sinusoids generates components at sums and differences of their frequencies. Violinists use the effect to check double stops: play two notes, listen for the difference tone, and its being in tune confirms the interval is.

What Roughness Does Not Explain

The roughness account is good. It is not complete, and three gaps are worth naming.

Roughness does not explain the pull of a resolution. A dominant seventh chord in tonal music demands to resolve, and the demand is far stronger than its modest roughness would justify. Musical expectation is learned, and it is doing work that acoustics does not describe.

Roughness is not the same as unpleasantness. Plenty of music is deliberately rough, and it is not thereby bad. Distorted guitar is almost pure roughness, and it is one of the most popular sounds of the last seventy years.

Roughness does not explain the octave. The octave has a special status, notes an octave apart are heard as "the same note", named with the same letter in essentially every musical culture, that goes well beyond its being smooth. A major sixth is nearly as smooth, and nobody calls it the same note.

The honest position is that the roughness model explains the *sensory* component of consonance well, and that musical consonance has a learned component on top of it that physics does not reach. This book will not pretend otherwise.

> **See Also**
> Musical Scales and Tuning Systems takes the ratios this chapter has justified and asks what happens when you try to build a scale out of them. The answer, that it cannot be done exactly, is the oldest unsolved problem in music.

8.6 Summary

- **Pitch resolution is proportional**: the just-noticeable difference is roughly 3–6 cents across the musical range, reflecting the cochlea's logarithmic map. It deteriorates above about 2 kHz and pitch perception largely fails above 5 kHz, where phase locking stops.
- **Beats** between two nearby tones occur at $|f_2 - f_1|$, since the sum is a tone at the average frequency modulated by an envelope at half the difference. Tuning by beats is far more precise than tuning by matching.
- **The missing fundamental**: a complex tone's pitch corresponds to the fundamental of its harmonic series, whether or not energy is present there. This is pattern matching,

not arithmetic on the lowest partial, mistuning one harmonic shifts the perceived pitch, and the partials can be split between the ears.

- **Place and temporal theories both contribute.** Timing dominates below 5 kHz and gives fine resolution and the missing fundamental; place takes over above, and performance degrades.
- **Roughness arises when two partials fall within one critical band** but are too far apart to beat slowly. For two *pure* tones, the roughest interval is about a semitone, and simple ratios have no special status.
- **For complex tones, simple ratios are smooth because their partials coincide.** Summing the roughness of every pair of partials produces a dissonance curve whose minima fall on the consonant intervals, derived from the ear alone, with no musical input.
- **Consonance is a relationship between a ratio and a timbre**, not a property of the ratio. Stretch the partials and the fifth turns rough; gamelan scales sit on the minima of their instruments' inharmonic partials.
- **Combination tones** at $f_2 - f_1$ and elsewhere arise from nonlinearity in the ear and are heard as separate pitches, not as beats.
- **The model has limits.** It does not explain harmonic expectation, it does not equate roughness with unpleasantness, and it does not explain the special status of the octave.

8.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple choice.** Pure tones at 440 Hz and 446 Hz sound together. Is the beat rate 3, 6, 443, or 886 beats per second?

2. **True or false.** A missing-fundamental pitch can be heard only when there is physical energy at the fundamental frequency.

3. **Drag the words.** Match beats, roughness, a combination tone, and a missing fundamental to their physical or perceptual origins.

4. **Fill in the blank.** Partials at 300, 400, and 500 Hz imply a missing fundamental of ____ Hz.

5. **Mark the words.** Identify intervals where harmonic complex tones tend to produce strong roughness minima: "Strong minima occur near the unison, octave, fifth, and fourth because many partials coincide; a semitone usually remains rough."

8.8 Conceptual Questions

1. Explain why the just-noticeable difference in pitch is nearly constant in cents but grows steeply in hertz, and relate this to the cochlea.

2. A piano tuner tunes a unison by listening for beats rather than by listening for sameness. Explain why this is more precise.

3. Explain how a telephone that transmits nothing below 300 Hz can convey a male voice whose fundamental is 110 Hz without the pitch sounding wrong.
4. State one piece of evidence that the missing fundamental is not computed simply as the difference between adjacent partials.
5. Explain why place theory struggles with the ear's pitch resolution, and why temporal theory struggles above 5 kHz.
6. Two *pure* tones a perfect fifth apart sound smooth, but so do two pure tones a tritone apart. Explain why simple ratios acquire their special status only for complex tones.
7. Explain why an interval that is consonant on a violin can be dissonant on a gamelan instrument, without appealing to cultural preference.
8. Give one musical phenomenon the roughness model does not explain, and say what kind of explanation it would need instead.

8.9 Problems

Exercise 8.1
(Straightforward) Two tones of 440 Hz and 444 Hz sound together. (a) What is the beat frequency? (b) What is the frequency of the tone that is heard? (c) How long between successive maxima?

Solution to Exercise 8.1
(a) $f_{\text{beat}} = |444 - 440| = 4$ Hz.
(b) The perceived tone is at the average:

$$\frac{440 + 444}{2} = 442 \text{ Hz}. \tag{8.8}$$

(c) $1/4 = 0.25$ s.
Therefore, a 442 Hz tone throbbing four times a second, with a quarter-second between peaks: an easily countable rate, and the working range for tuning.

Exercise 8.2
(Moderate) A guitarist tunes the B string against the fretted G string. They hear 3 beats per second and the fretted note is 246.9 Hz. (a) What are the two possible frequencies of the B string? (b) The guitarist tightens it slightly and the beats speed up to 5 per second. Which was it?

Solution to Exercise 8.2
(a) Three beats per second means the two differ by 3 Hz:

$$246.9 + 3 = 249.9 \text{ Hz} \quad \text{or} \quad 246.9 - 3 = 243.9 \text{ Hz}. \tag{8.9}$$

(b) Tightening raises the pitch. If the string had been *flat* at 243.9 Hz, tightening would move it toward 246.9 and the beats would slow. They sped up, so the string was already **sharp**, at 249.9 Hz.
Therefore, the string was at 249.9 Hz and must be loosened. This "tighten and see which way the beats go" test is the standard way to resolve the ambiguity, and it is exactly why the beat method works when simple comparison does not.

Exercise 8.3
(Straightforward) A complex tone contains partials at 440, 660, 880, and 1100 Hz. (a) What is the fundamental? (b) Which harmonics are these? (c) What pitch is heard?

Solution to Exercise 8.3
(a) The partials are spaced 220 Hz apart and all are multiples of 220 Hz, so $f_1 = 220$ Hz.
(b) $440 = 2f_1$, $660 = 3f_1$, $880 = 4f_1$, $1100 = 5f_1$: the 2nd through 5th harmonics.
(c) 220 Hz.
Therefore, the pitch heard is A3 at 220 Hz, an octave below the lowest partial actually present. Note that 440 Hz alone would also be consistent with partials at 440 and 880, but not with 660 and 1100; so the pattern-matching account settles on 220.

Exercise 8.4
(Moderate) A tuner sets a tempered major third above $\mathrm{C}_4 = 261.6$ Hz. The equal-tempered third is 400 cents; the just third is 5:4. (a) Find both frequencies. (b) Find the beat rate between the 5th harmonic of the lower note and the 4th harmonic of the tempered upper note.

Solution to Exercise 8.4
(a) Equal-tempered:

$$f = 261.6 \times 2^{400/1200} = 261.6 \times 1.259921 = 329.6 \text{ Hz}. \tag{8.10}$$

Just:

$$f = 261.6 \times \frac{5}{4} = 327.0 \text{ Hz}. \tag{8.11}$$

(b) The relevant harmonics:

$$5 \times 261.6 = 1308.0 \text{ Hz}, \qquad 4 \times 329.6 = 1318.4 \text{ Hz}, \tag{8.12}$$

$$f_{\text{beat}} = 10.4 \text{ Hz}. \tag{8.13}$$

Therefore, the tempered major third beats about ten times a second, fast enough to be heard as a distinct shimmer rather than counted. The equal-tempered third is therefore the interval musicians most often complain about, and Chapter 9's meantone temperaments were willing to sacrifice so much to improve it.

Exercise 8.5
(Straightforward) The just-noticeable difference near 440 Hz is about 4 cents. (a) What frequency difference is that in hertz? (b) How does it compare with the difference between the just and equal-tempered major thirds from Exercise 8.4?

Solution to Exercise 8.5
(a) Four cents is a ratio of $2^{4/1200} = 1.00231$, so

$$\Delta f = 440 \times 0.00231 = 1.0 \text{ Hz}. \tag{8.14}$$

(b) The just third is 327.0 Hz and the tempered one 329.6 Hz, a difference of

$$1200 \log_2(329.6/327.0) = 14 \text{ cents}. \tag{8.15}$$

Therefore, the discrepancy between the two thirds is about three and a half times the threshold of detection, comfortably audible, and precisely the reason tuning systems are argued about.

Exercise 8.6

(Moderate) Two tones at 1000 Hz and 1080 Hz are played together. The critical bandwidth at 1000 Hz is about 160 Hz. (a) Do they fall in the same critical band? (b) Predict what is heard. (c) Repeat for tones at 1000 and 1400 Hz.

Solution to Exercise 8.6

(a) They differ by 80 Hz, which is half the critical bandwidth; so yes, comfortably within one band.

(b) 80 Hz is far too fast to count as beats and well inside a critical band: the prediction is **roughness**, a harsh buzzing quality, and the two are not heard as separate tones.

(c) 400 Hz apart, which is two and a half critical bandwidths. They fall in separate bands, so the prediction is **two smooth, separate tones** with no roughness.

Therefore, the same pair of pure tones can be rough or smooth depending only on their separation relative to the critical band, and 400 Hz, a major sixth, is smooth not because of its ratio but because it is wide.

Exercise 8.7

(Moderate) Two complex tones with harmonic partials sound a perfect fifth apart, the lower at 200 Hz. (a) List the first six partials of each. (b) Identify every coincidence. (c) Explain what the coincidences have to do with consonance.

Solution to Exercise 8.7

(a) Lower (200 Hz): 200, 400, 600, 800, 1000, 1200.

Upper (300 Hz): 300, 600, 900, 1200, 1500, 1800.

(b) Coincidences: 600 Hz (lower's 3rd with upper's 2nd) and 1200 Hz (lower's 6th with upper's 4th).

(c) A coinciding pair of partials contributes **zero** roughness, since roughness requires a small non-zero separation. Every coincidence therefore removes a pair that would otherwise have been a candidate for clashing. Detuning the fifth slightly turns each exact coincidence into a near-miss a few hertz wide, precisely the most rough condition, so the dissonance rises steeply on either side of the exact ratio.

Therefore, the fifth's consonance comes from its partials lining up, and the sharpness of the minimum comes from how quickly that alignment is lost.

Exercise 8.8

(Moderate) An organ has an 8-foot stop sounding 65.4 Hz and a $5\frac{1}{3}$-foot stop sounding a fifth above. (a) What is the frequency of the second? (b) What difference tone results? (c) What pitch does the listener report, and what is the stop called?

Solution to Exercise 8.8

(a) A fifth above is a ratio of 3/2:

$$f = 65.4 \times 1.5 = 98.1 \text{ Hz}. \tag{8.16}$$

(b) The difference tone is

$$98.1 - 65.4 = 32.7 \text{ Hz}. \tag{8.17}$$

(c) 32.7 Hz is exactly an octave below 65.4 Hz. The listener reports a note an octave below the 8-foot stop, a 16-foot pitch, and the arrangement is called a **resultant** or **acoustic bass**.

Therefore, a builder can obtain a 16-foot pitch without a 16-foot pipe. The effect works partly through the difference tone and, more robustly, through the missing-fundamental mechanism of §8.3: 65.4 and 98.1 Hz are the 2nd and 3rd harmonics of 32.7 Hz, and the ear supplies the fundamental. It is a common solution where a church has neither the height nor the budget for the real thing.

Exercise 8.9

(Challenging) Using the beat formula, show that two tones at f and $f+\Delta$ produce an envelope at $\Delta/2$ but a beat rate of Δ. Explain the factor of two.

Solution to Exercise 8.9

Applying the sum-to-product identity:

$$\sin(2\pi f t) + \sin\big(2\pi(f+\Delta)t\big) = 2\cos\left(2\pi\frac{\Delta}{2}t\right)\sin\left(2\pi\left(f+\frac{\Delta}{2}\right)t\right). \tag{8.18}$$

The modulating factor is $\cos(2\pi(\Delta/2)t)$, which completes $\Delta/2$ cycles per second.

But the ear responds to the *magnitude* of the envelope, not its sign. The function $|\cos|$ has twice the frequency of cos, because it reaches a maximum both when the cosine is $+1$ and when it is -1. So loudness peaks occur $2 \times (\Delta/2) = \Delta$ times per second.

Therefore, the envelope oscillates at $\Delta/2$ and the audible beat rate is Δ. The factor of two comes entirely from the ear's indifference to the sign of the envelope, and forgetting it is the classic error in this calculation.

Exercise 8.10

(Moderate) The critical bandwidth is roughly a third of an octave. Two complex tones are played a minor second apart (100 cents). (a) How does this compare with a critical band? (b) Predict the roughness. (c) Repeat for two tones a minor ninth apart (1300 cents).

Solution to Exercise 8.10

(a) A third of an octave is 400 cents, so a minor second at 100 cents is a quarter of a critical band.

(b) A quarter of a critical band is very close to where Figure 8.8 puts the roughness maximum. The prediction is maximum roughness, and the fundamentals are only the start, since every pair of corresponding harmonics is *also* a quarter of a critical band apart in musical terms and contributes its own roughness.

(c) A minor ninth is a minor second plus an octave. The two fundamentals are now well over a critical band apart and contribute nothing. But the *upper* partials still clash: the lower tone's 2nd harmonic is a minor second below the upper tone's fundamental, and so on up.

Therefore, the minor ninth is smoother than the minor second but by no means smooth, which matches musical practice, where the minor ninth is used as a pungent dissonance rather than as a consonance, and composers routinely spread a dissonant interval across an octave to soften it without removing it.

Exercise 8.11

(Moderate) A synthesizer produces tones whose partials lie at $n^{1.1}$ times the fundamental rather than n times. (a) Find the first four partials of a 200 Hz tone. (b) Two such tones are played a perfect fifth apart. Do any partials coincide? (c) Comment on the consonance.

Solution to Exercise 8.11

(a) Partials at $200 \times n^{1.1}$:

$$n = 1: \ 200, \quad n = 2: \ 429, \quad n = 3: \ 669, \quad n = 4: \ 920 \text{ Hz}. \tag{8.19}$$

(b) The upper tone at 300 Hz has partials 300, 643, 1003, 1379 Hz. Comparing with the lower's 200, 429, 669, 920: the closest approach is 669 against 643, which differ by 26 Hz. Nothing coincides.

(c) With no coincidences, and with several pairs separated by a few tens of hertz, squarely inside a critical band at these frequencies, the fifth will sound rough rather than settled.

Therefore, a 3:2 ratio is not consonant on this timbre. The consonant intervals for these partials would be found by computing the dissonance curve, and they would not be the intervals of the Western scale. This is the Sethares argument of §8.4, made concrete.

Exercise 8.12

(Challenging) Two violinists play a double stop. The lower note is 392 Hz and the upper is meant to be a just major sixth above. (a) What should the upper frequency be? (b) What difference tone results? (c) The upper player is 8 cents sharp. Recompute the difference tone and comment on whether the error is easier to hear directly or through the difference tone.

Solution to Exercise 8.12

(a) A just major sixth is 5:3:

$$f = 392 \times \frac{5}{3} = 653.3 \text{ Hz}. \tag{8.20}$$

(b) The difference tone:

$$653.3 - 392 = 261.3 \text{ Hz}, \tag{8.21}$$

which is a perfect fifth below the lower note: a musically consonant addition.

(c) Eight cents sharp:

$$f' = 653.3 \times 2^{8/1200} = 653.3 \times 1.00463 = 656.4 \text{ Hz}, \tag{8.22}$$

$$f'_{\text{diff}} = 656.4 - 392 = 264.4 \text{ Hz}. \tag{8.23}$$

The difference tone has moved by 3.1 Hz, which is $1200 \log_2(264.4/261.3) = 20$ cents.

Therefore, an 8 cent error in the upper note produces a 20 cent shift in the difference tone: the error is **amplified by a factor of 2.5**. This is exactly why string players listen for difference tones when tuning double stops: the mechanism turns a marginal error into an obvious one.

Exercise 8.13

(Challenging) A listener is presented with the odd harmonics of 100 Hz in the left ear and the even harmonics in the right. (a) What pitch do they report? (b) Explain why this result is difficult for any theory placing pitch extraction in the cochlea.

Solution to Exercise 8.13

(a) 100 Hz: the fundamental of the combined series.

(b) Neither cochlea receives a complete harmonic series. The left receives 100, 300, 500, …which is consistent with a fundamental of 100 Hz, but the right receives 200, 400, 600, …which on its own is a perfectly good harmonic series on 200 Hz. If pitch were extracted independently in each cochlea and then combined, the listener should hear two pitches an octave apart, or should hear the right ear's 200 Hz.

They do not. They hear a single 100 Hz pitch, which requires the partials from both ears to have been pooled *before* the fundamental was determined.

Therefore, pitch extraction must occur centrally, at or above the first binaural stage of the auditory pathway. No purely cochlear mechanism, place or temporal, can produce this result.

Exercise 8.14

(Challenging) A gamelan metallophone has partials at 1, 2.76, 5.40, and 8.93 times its fundamental. (a) Two such instruments play an interval of 1.5 (a Western fifth). Find the partials of each and identify any near-coincidences. (b) Suggest, qualitatively, what kind of interval *would* produce coincidences for this timbre.

Solution to Exercise 8.14

(a) Taking the lower fundamental as 1:

Lower: 1, 2.76, 5.40, 8.93.

Upper (at 1.5): 1.5, 4.14, 8.10, 13.40.

Comparing: 1.5 against nothing nearby; 4.14 against 5.40 (a ratio of 1.30, far apart); 8.10 against 8.93 (a ratio of 1.10, about 170 cents, close enough to be within a critical band at these frequencies, and therefore *rough*).

So not only are there no coincidences, there is at least one actively clashing pair.

(b) Coincidences would require an interval r such that r times one of the lower partials lands on another. For instance, $r = 2.76$ would put the upper fundamental on the lower's second partial: an interval of $1200 \log_2(2.76) = 1758$ cents, about an octave and a fourth. Ratios built from the numbers 2.76, 5.40 and 8.93 and their quotients are the candidates.

Therefore, the intervals that are consonant on this instrument are not the Western ones, and a scale built for it would not be the Western scale. Which is, of course, what gamelan tuning actually is.

Part IV

Scales and Tuning

Chapter 9

Musical Scales and Tuning Systems

Learning Objectives

By the end of this chapter, you should be able to:

- Express a musical interval as a frequency ratio, and explain why intervals multiply where pitches add.
- Define the cent, convert between a frequency ratio and an interval in cents, and use cents to compare tunings that differ by only a few parts per thousand.
- Construct the Pythagorean scale from a chain of pure fifths, and calculate the Pythagorean comma that results from twelve of them.
- Explain why no stack of pure fifths ever closes into an octave, as a consequence of the fact that no power of 3/2 is a power of 2.
- Construct a just major scale from the ratios 3:2, 5:4, and 6:5, and identify the syntonic comma and the two sizes of whole tone it produces.
- Explain concretely why a keyboard tuned in just intonation cannot modulate, using the wolf fifth as the example.
- Describe quarter-comma meantone and the well temperaments as deliberate distributions of the comma, and state what each was optimizing for.
- Derive equal temperament from the requirement that twelve equal semitones make an octave, and calculate the ratio of the equal-tempered semitone.
- Compare equal-tempered intervals with their just counterparts in cents, and identify which are nearly exact and which are noticeably wide or narrow.
- Explain why fretted and keyboard instruments are bound to a fixed tuning while singers, string quartets, and trombonists are not.

Introduction

Chapter 8 ended with a result worth restating: intervals whose frequency ratios are simple whole numbers sound smooth, because the partials of the two tones coincide instead of clashing. The octave is 2:1, the perfect fifth is 3:2, the major third is 5:4.

So building a musical scale ought to be easy. Take the good ratios and stack them up.

It cannot be done. Not approximately, not with more effort, not with better arithmetic, **the good ratios do not fit into an octave**, and no amount of cleverness will make them. This is an arithmetical fact, not a historical accident or a limitation of any instrument, and every tuning system ever devised is a different answer to the question of what to do about it.

The obstruction is easy to state. Stack twelve perfect fifths and you get *almost* seven



Figure 9.1: Instrument makers have repeatedly reshaped resonators to change how partials reach the listener. Historical image, public domain.

octaves, missing by about a quarter of a semitone. Stack four perfect fifths and you get *almost* a major third plus two octaves, missing by about a fifth of a semitone. Those two small discrepancies, the **Pythagorean comma** and the **syntonic comma**, are the entire subject of this chapter. Every tuning system is a decision about where to put them.

That decision has consequences musicians can hear, and the history of Western music is bound up with it. The reason Bach could write in all twenty-four keys, and his grandfather's generation could not, is a change in where the comma was hidden.

9.1 Intervals as Ratios

The Octave and the Fifth

An interval is a **ratio** of frequencies, not a difference. The octave above 220 Hz is 440 Hz, and the octave above 440 Hz is 880 Hz. The differences are 220 and 440 Hz; the ratio is 2 in both cases, and it is the ratio the ear reports.

This follows directly from Chapter 6: the cochlea's frequency map is logarithmic, so equal ratios occupy equal distances along it, and Chapter 8 showed that pitch discrimination is likewise constant in proportional terms.

The intervals that matter most are the first few of the harmonic series:

Interval	Ratio	Harmonics involved
Octave	2:1	1 and 2
Perfect fifth	3:2	2 and 3
Perfect fourth	4:3	3 and 4
Major third	5:4	4 and 5
Minor third	6:5	5 and 6

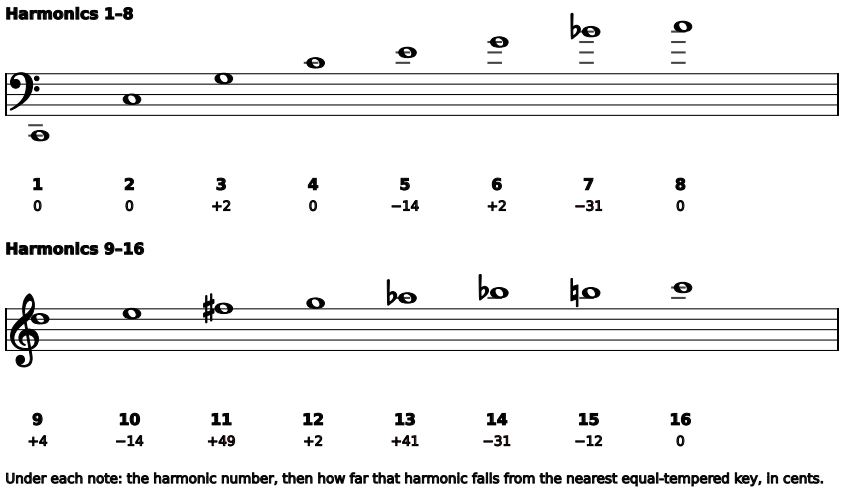


Figure 9.2: The harmonic series again, now read as a source of intervals. The first six partials give the octave, the fifth, the fourth, and the major third, the consonances of Western music, in order. From the seventh partial upward the series diverges sharply from anything a keyboard can play, which is why those partials play almost no part in common-practice harmony.

Intervals Multiply

Stack two intervals and the ratios **multiply**:

$$\text{fifth} \times \text{fourth} = \frac{3}{2} \times \frac{4}{3} = 2 = \text{octave}. \tag{9.1}$$

This is why a fifth plus a fourth is an octave, and it is also the source of a persistent confusion. Musicians count intervals by *adding*, a third plus a third is a fifth, because they are counting note names, which is a logarithmic count. The physics multiplies. The two are consistent, and cents are what reconciles them.

The Cent

The **cent** is one hundredth of an equal-tempered semitone, so there are 1200 to the octave. Converting between a ratio r and an interval n in cents:

$$n = 1200 \log_2 r, \qquad r = 2^{n/1200}. \tag{9.2}$$

Cents are logarithmic, so they **add** the way musicians expect while representing ratios exactly. And the unit is well matched to the ear: Chapter 8 put the just-noticeable difference at 3–6 cents, so an error of a cent or two is inaudible and one of fifteen is obvious.

The essential values, worth memorizing:

Interval	Ratio	Cents
Octave	2:1	1200.0
Pure fifth	3:2	701.955
Pure fourth	4:3	498.045
Pure major third	5:4	386.314
Pure minor third	6:5	315.641
Equal semitone	$2^{1/12}$	100.0

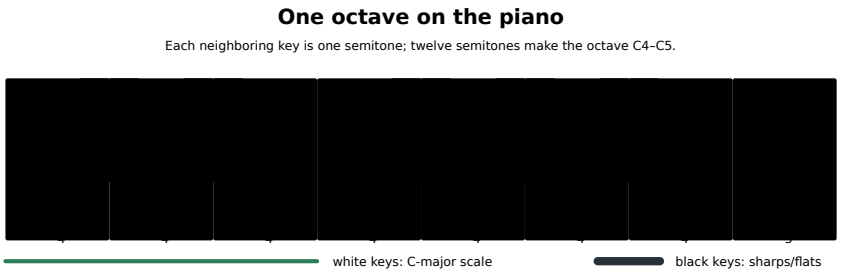


Figure 9.3: Reading notes on a keyboard. The white keys trace the C-major scale, while each black key has two names: for example, C-sharp and D-flat are the same equal-tempered key. Moving to the next key, white or black, raises the pitch by one semitone; the octave from C4 to C5 contains twelve such steps.

Worked example: converting to and from cents

(a) How many cents is the ratio 7:4*, the interval of the seventh harmonic? (b) What ratio is 350 cents?*

(a)

$$n = 1200 \log_2(7/4) = 1200 \log_2(1.75) = 1200(0.80735) = 968.8 \text{ cents}. \tag{9.3}$$

The nearest equal-tempered interval is the minor seventh at 1000 cents, so the seventh harmonic is 31 cents flat of it, as the staff figure above says.

(b)

$$r = 2^{350/1200} = 2^{0.29167} = 1.2240. \tag{9.4}$$

That lies between the pure minor third (1.2) and the pure major third (1.25): a "neutral third", which appears in Arabic and Turkish music and in no Western keyboard.

9.2 Pythagorean Tuning

A Chain of Pure Fifths

The oldest systematic tuning takes the two simplest ratios, the octave 2:1 and the fifth 3:2, and uses nothing else.

Start on a note and go up by fifths: C, G, D, A, E, B, F♯, …Each new note is 3/2 times the last; whenever the pitch leaves the octave, halve it to bring it back. After twelve steps every chromatic note has been generated.

This produces beautiful fifths, by construction: every fifth in the chain is exactly pure. The fourths are pure too, being octave complements of fifths.

The Pythagorean Comma

Now the problem. Twelve fifths should return to the starting note, seven octaves higher. They do not.

$$\left(\frac{3}{2}\right)^{12} = 129.746, \qquad 2^7 = 128. \tag{9.5}$$

The ratio between them is

$$\frac{129.746}{128} = 1.013643, \tag{9.6}$$

which is

$$1200 \log_2(1.013643) = 23.46 \text{ cents}. \tag{9.7}$$

This is the **Pythagorean comma**, and it is not going away. The reason is elementary number theory: $(3/2)^{12} = 3^{12}/2^{12}$, and no power of 3 is ever a power of 2, since 3 and 2 are distinct primes. The chain of fifths can *never* close, at twelve steps or at any other number.

The Scale It Produces

Pythagorean tuning gives pure fifths and fourths, and, as the price, a major third of

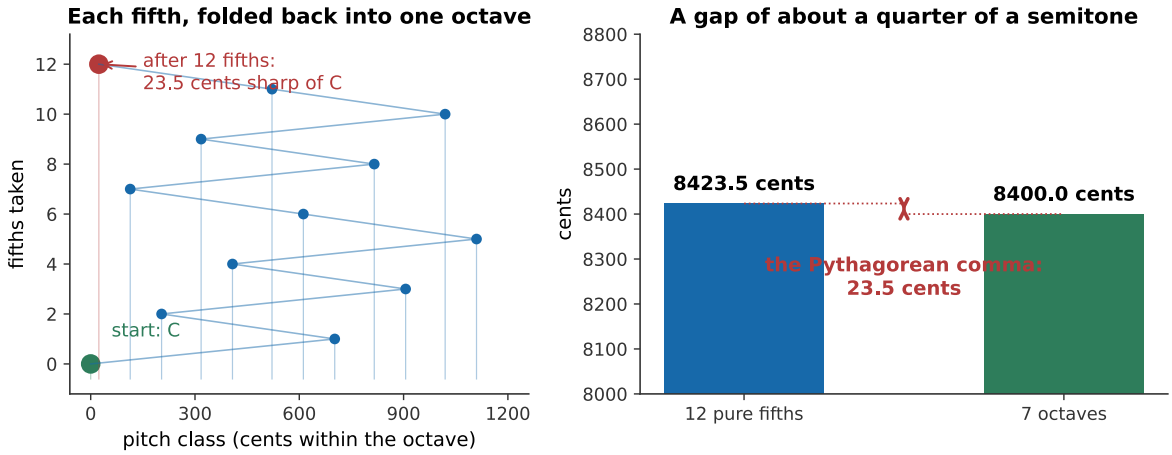


Figure 9.4: The chain of fifths fails to close. **Left**: each fifth folded back into a single octave; the twelfth lands not on C but 23.5 cents above it. **Right**: the same fact as arithmetic. Twelve pure fifths overshoot seven octaves by the Pythagorean comma.

$$\left(\frac{3}{2}\right)^4 \div 4 = \frac{81}{64} = 407.8 \text{ cents}, \tag{9.8}$$

against a pure 5:4 of 386.3 cents. That is 21.5 cents sharp, an error known as the **syntonic comma**.

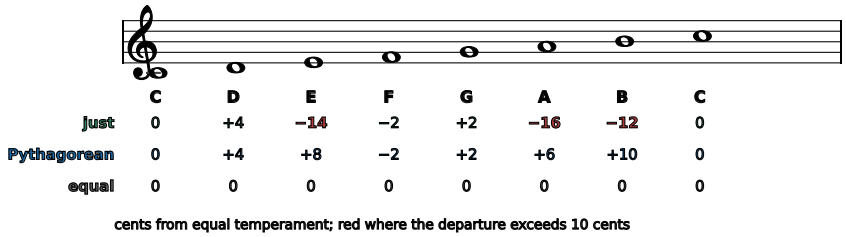


Figure 9.5: The same seven notes, tuned three ways. Every system agrees on C, F, and G to within a couple of cents; they disagree about the thirds and sixths, and they disagree in opposite directions. Pythagorean thirds are sharp, just thirds are flat of equal temperament, and the gap between them is the syntonic comma.

Twenty-one cents is four times the threshold of detection, and on sustained tones it beats audibly. For medieval music this was tolerable, because the third was treated as a *dissonance* requiring resolution: the consonances were the octave, fifth, and fourth, which Pythagorean tuning renders perfectly. When the third became a consonance in the fifteenth century, the tuning had to change.

9.3 Just Intonation

Bringing In the Fifth Harmonic

Just intonation adds the ratio 5:4 to the toolkit and builds the scale from three pure triads: on the tonic, the subdominant, and the dominant. Each triad is tuned 4:5:6: a pure major third and a pure fifth.

The resulting C major scale:

Note	Ratio	Cents
C	1:1	0
D	9:8	203.9
E	5:4	386.3
F	4:3	498.0
G	3:2	702.0
A	5:3	884.4
B	15:8	1088.3
C	2:1	1200

Played in C, this is gorgeous. The tonic, subdominant, and dominant triads are all exactly pure, with every partial coinciding and nothing beating at all.

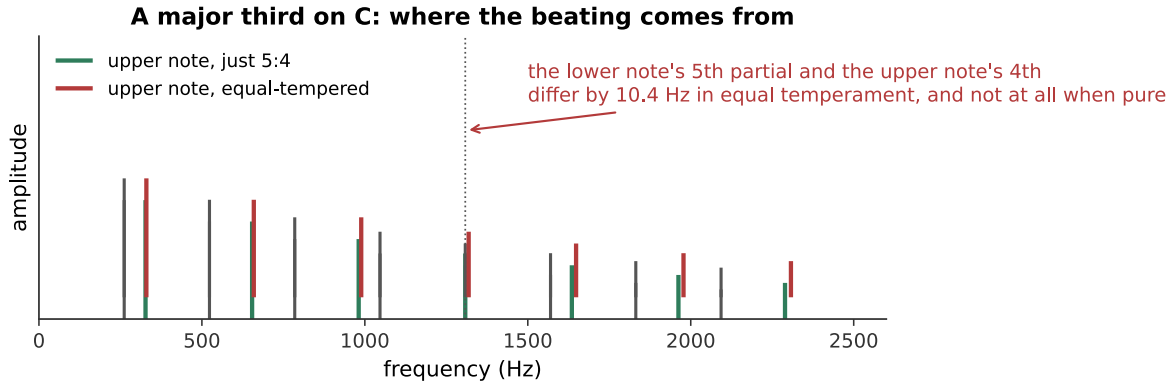


Figure 9.6: A C major triad, tuned justly and in equal temperament. The difference is not one of pitch, both are unmistakably C major, but of *stillness*. The just triad has nothing beating in it; the tempered one beats about ten times a second in its third. *What you hear:*Two C major triads. The first is still and glassy, with no movement in it at all. The second has an audible shimmer, a slow waver that the first does not have. Listen: Just, Equal-tempered

Two Sizes of Whole Tone, and the Wolf

The trouble appears as soon as you look at the intervals *between* the scale degrees.

$$\text{C to D} = \frac{9}{8} = 203.9 \text{ cents}, \qquad \text{D to E} = \frac{5/4}{9/8} = \frac{10}{9} = 182.4 \text{ cents}. \tag{9.9}$$

There are **two different whole tones**, differing by 21.5 cents: the syntonic comma again. A keyboard has one key for each, so it must choose.

The consequence is a fifth that is badly out of tune:

$$\text{D to A} = \frac{5/3}{9/8} = \frac{40}{27} = 680.4 \text{ cents}, \tag{9.10}$$

against a pure 702 cents. That fifth is 21.5 cents narrow, and it beats violently. It is the **wolf fifth**, named for its howl.

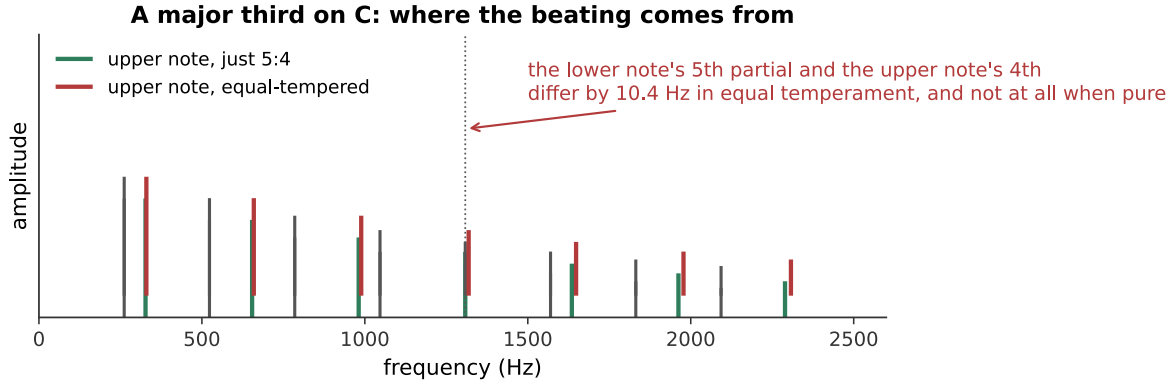


Figure 9.7: The major third alone, where the effect is clearest. The figure shows why: the lower note's fifth partial and the upper note's fourth partial coincide exactly when the third is pure, and differ by about 10 Hz when it is tempered. *What you hear:*Two major thirds. The first is perfectly steady; the second has a distinct and rather fast beating in it, about ten beats a second. Listen: Just 5:4, Equal-tempered

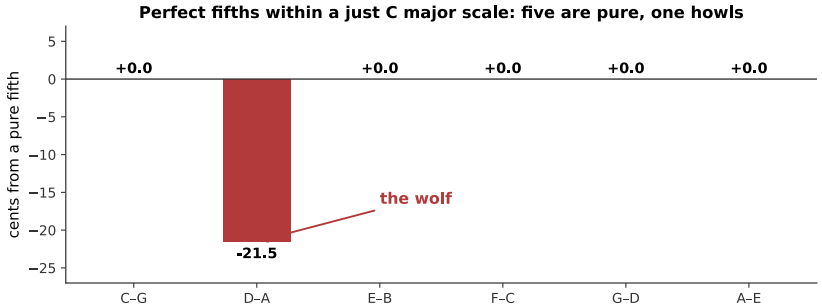


Figure 9.8: The fifths inside a just C major scale. Five are exactly pure. One, D–A, is a syntonic comma narrow, and it is unusable in sustained harmony.

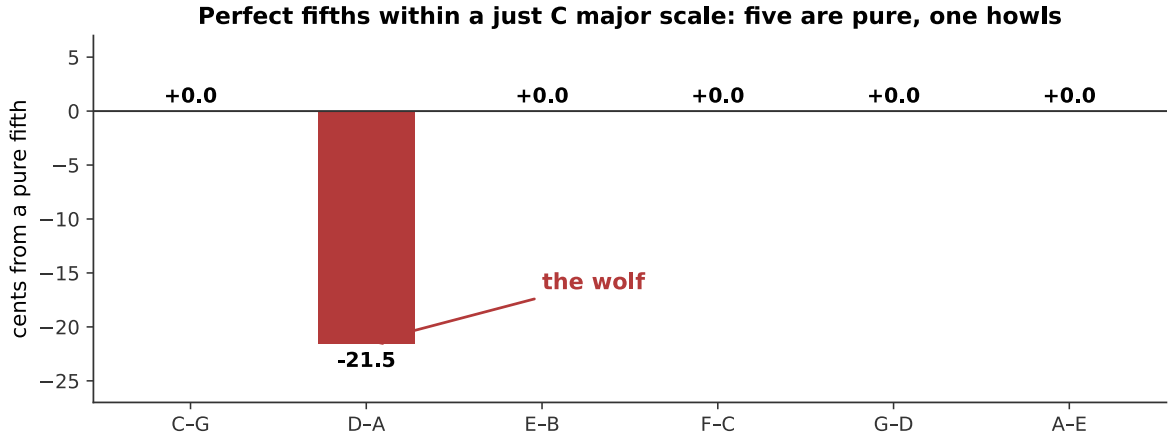


Figure 9.9: A pure fifth and the wolf. The wolf is not subtly out of tune; it is the kind of thing a listener assumes is a mistake. *What you hear:*Two fifths. The first is perfectly still. The second has a fast, insistent beating that makes it sound actively wrong rather than merely different. Listen: A pure fifth, The wolf fifth

Why Just Intonation Cannot Modulate

The deeper problem is not the single wolf. It is that just intonation is tuned **for one key**, and there is no way to make a fixed set of twelve pitches serve them all.

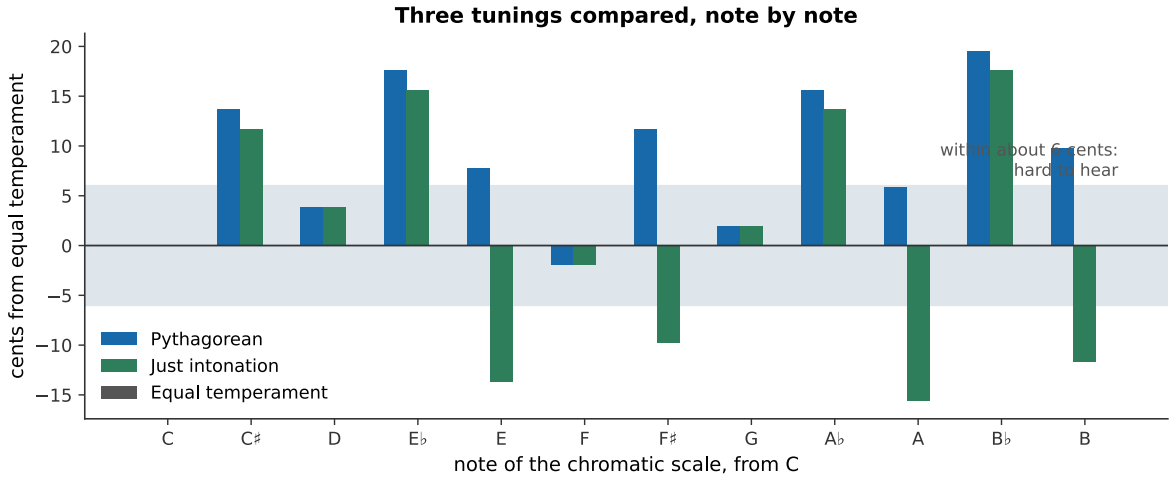


Figure 9.10: One set of twelve pitches, tuned justly for C, playing the same progression in two keys. The instrument has not been retuned between the two; the *music* has moved, and the tuning has not moved with it. This is the practical argument that ended just intonation as a keyboard tuning. *What you hear:*The same four-chord progression, played twice on the same twelve fixed pitches. In C it is clean and settled. Moved up a tone to D, the same chords are audibly sour, with obvious beating. Listen: I–IV–V–I in C, The same progression in D

Singers, string quartets, and trombonists do not have this problem, because they adjust continuously. They routinely play closer to just intonation than to equal temperament: a good choir will tune a sustained chord pure without being asked. **Just intonation was never abandoned by musicians who can bend their pitch. It was abandoned by keyboards**, and only because a keyboard has to decide in advance.

9.4 Meantone and the Well Temperaments

Spending the Comma Deliberately

The insight that unlocked the problem is that the comma has to go *somewhere*, but it does not have to go all in one place. Distribute it and no single interval is ruined.

A system that does this deliberately is a **temperament**, and the choice of how to distribute is a choice about which keys will sound best.

Quarter-Comma Meantone

The most important historical temperament narrows every fifth by a quarter of a syntonic comma, about 5.4 cents, which is just at the edge of audibility.

Four narrowed fifths then produce a major third that is *exactly* pure:

$$4 \times (701.955 - 5.377) = 2786.3 \text{ cents}, \tag{9.11}$$

which, less two octaves, is 386.3 cents: the pure 5:4.

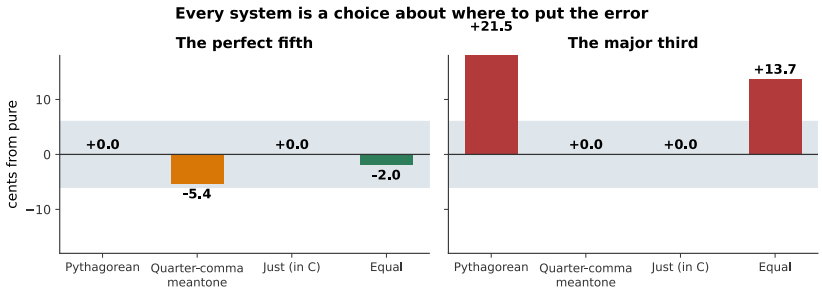


Figure 9.11: Where each system puts its error. Pythagorean tuning buys pure fifths at the cost of a 22 cent third. Quarter-comma meantone buys pure thirds at the cost of a 5 cent fifth, which is barely audible. Equal temperament splits the difference, with a fifth 2 cents narrow and a third 14 cents sharp.

Meantone dominated European keyboard music for two centuries, and for good reason: it makes major thirds beatless, and thirds are what Renaissance and early Baroque harmony is built on. Its price is that the twelve fifths still cannot close, so the accumulated error is dumped into one interval, traditionally G♯–E♭, which becomes a wolf of about 737 cents, 35 cents wide and completely unusable. Composers simply avoided the keys that needed it.

Well Temperament and Key Color

By the late seventeenth century composers wanted more keys, and **well temperaments** were the answer: distribute the comma unevenly, so that no interval is unusable but the distribution is not uniform either.

The consequence is that **each key sounds different**. Keys with few accidentals are nearly pure; remote keys are progressively more tempered and sound more restless. This is "key color", and eighteenth-century writers described keys in terms that sound like synesthesia, D major brilliant, E♭ major solemn, F minor mournful, and were describing something real about their instruments.

> **Note**
> Bach's *Das wohltemperirte Clavier* means "the well-tempered keyboard", not "the equal-tempered keyboard", and the distinction matters. The collection demonstrates that all twenty-four keys are *usable*, which is a well-temperament claim. In equal temperament all keys are not merely usable but identical, and the point of writing in all of them would largely vanish.

9.5 Equal Temperament

Twelve Equal Semitones

Equal temperament takes the distribution to its limit: divide the octave into twelve exactly equal steps.

The requirement is that twelve semitones make an octave, so each semitone is

$$r = 2^{1/12} = 1.059463, \tag{9.12}$$

which is 100 cents by definition. Every interval of n semitones is $2^{n/12}$.

The mathematics is forced. Twelve equal multiplicative steps making a factor of two admits exactly one solution, and the twelfth root of two is irrational; so **no** equal-tempered interval except the octave is ever a whole-number ratio.

Worked example: the whole keyboard from one number
Given $A_4 = 440$ Hz, find the frequency of C_5 (three semitones up) and of A_2 (two octaves down).

$$f_{C_5} = (440\text{ Hz}) \times 2^{3/12} = (440\text{ Hz})(1.18921) = 523.25\text{ Hz}. \tag{9.13}$$

$$f_{A_2} = (440\text{ Hz}) \times 2^{-24/12} = \frac{440\text{ Hz}}{4} = 110\text{ Hz}. \tag{9.14}$$

Every note on a modern keyboard follows from $f = 440 \times 2^{n/12}$, with n the number of semitones from A_4. That single formula replaced two thousand years of argument about ratios.

What It Costs and What It Buys

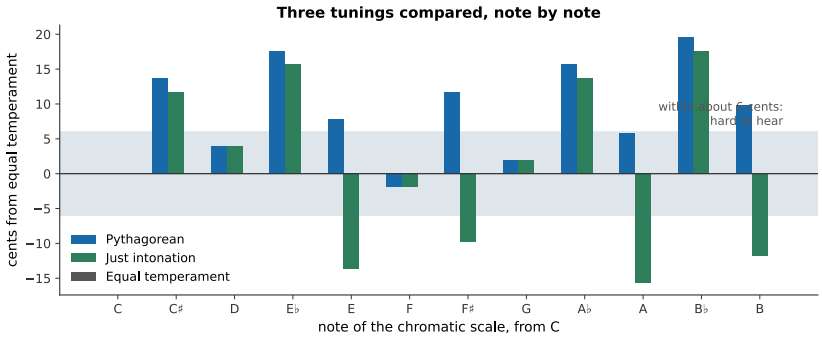


Figure 9.12: Three tunings, note by note. Equal temperament is the zero line by construction. Pythagorean and just intonation disagree with it, and with each other, by up to about 20 cents, mostly on the thirds and sixths.

The accounting for equal temperament:

Interval	Equal	Pure	Error
Octave	1200	1200	0
Fifth	700	701.96	-1.96
Fourth	500	498.04	+1.96
Major third	400	386.31	+13.69
Minor third	300	315.64	-15.64
Major sixth	900	884.36	+15.64

The fifth is 2 cents narrow, well below the threshold of detection, and essentially perfect. The thirds and sixths are 14 to 16 cents out, which is clearly audible on sustained tones.

So equal temperament is a *worse* compromise than meantone for any single key. What it buys is that **the compromise is the same everywhere**. Every key is equally usable, unlimited modulation is possible, and any piece can be transposed without retuning.

The nineteenth century decided that was worth an audibly imperfect third, and the decision has held. Notice that this was a trade of *quality* for *freedom*, made by a musical culture that had begun to care more about harmonic adventure than about the purity of a sustained chord.

Comparing the Systems in Cents

A summary of the three main systems in one table:

	Fifths	Major thirds	Modulation
Pythagorean	All pure	22 cents sharp	Impossible (comma)
Just	Mostly pure, one wolf	Pure in home key	Impossible (wrong key)
Quarter-comma meantone	5 cents narrow	Pure	Limited (wolf at G♯–E♭)
Equal	2 cents narrow	14 cents sharp	Unlimited

Read down the last column and the history explains itself.

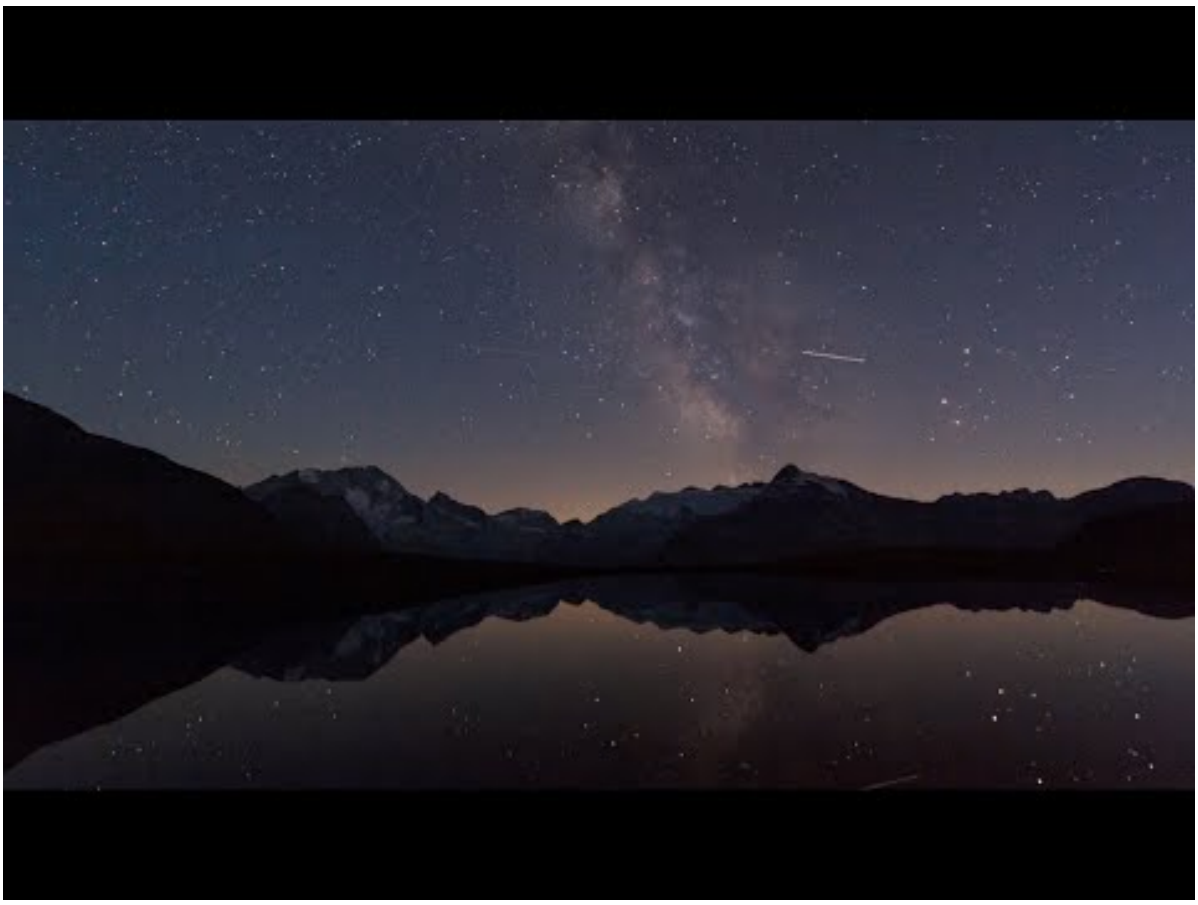


Figure 9.13: Adam Neely compares equal temperament with just intonation from a working musician's perspective. The examples make audible both the pure intervals that just tuning gains and the freedom to change key that equal temperament buys. Watch on YouTube: Equal Temperament vs. Just Intonation

9.6 Tuning in Practice

Fixed-Pitch and Flexible-Pitch Instruments

Equal temperament is universal on **fixed-pitch** instruments, piano, organ, harp, fretted strings, mallet percussion, because those instruments must commit in advance.

It is much less universal among **flexible-pitch** instruments, which adjust as they play:

- A **string quartet** tunes sustained chords toward just intonation, narrowing thirds instinctively.
- A **choir** does the same, more strongly; a well-tuned a cappella chord is close to pure.
- A **trombonist** has a continuously variable slide and uses it.
- **Wind players** adjust with embouchure and alternate fingerings continuously.

Ensembles mixing the two families negotiate. A string quartet plays closer to just; add a piano and the strings bend toward it, because the piano cannot bend toward them.

> **Warning**
> It is tempting to say flexible-pitch players "play in just intonation". They do not, quite. They play *expressively*, which often means the opposite: a leading note is frequently played **sharp**, sharper even than equal temperament, to intensify its pull toward the tonic. That is a melodic impulse pulling against the harmonic one, and good players trade between them constantly.

Stretched Tuning on the Piano

One place where even equal temperament is not quite applied is the piano, for the reason Chapter 5 gave.

Piano strings are stiff, so their partials are **inharmonic**: the nth partial is sharp of n times the fundamental. A tuner setting an octave listens to the lower note's second partial against the upper note's fundamental, and since that second partial is sharp, the octave must be stretched to match it.

The result is the **Railsback curve**: a piano's treble is tuned progressively sharp and its bass progressively flat, by as much as 30 cents at the extremes. A piano tuned to a mathematically exact equal temperament sounds *out of tune*, and every tuner stretches without needing to be told to. Chapter 10 works through the mechanism.

Scales Outside the Western Tradition

Twelve equal divisions is one solution among many, and the alternatives are not approximations to it.

Indian classical music uses *shrutis*, traditionally twenty-two per octave, and raga performance draws on just ratios with ornaments that pass between them. The tradition has no fixed-pitch keyboard to force a compromise, so it never needed one.

Arabic and Turkish music use intervals near the quarter tone, including the neutral third of about 350 cents computed in §9.1: an interval with no equal-tempered equivalent at all.

Javanese and Balinese gamelan use *slendro* (roughly five near-equal steps) and *pelog* (seven unequal steps), and no two gamelan sets are tuned alike. Chapter 8 gave the reason this

is coherent rather than arbitrary: gamelan instruments have **inharmonic** partials, so their dissonance curve has its minima somewhere other than the simple ratios, and the scales sit on those minima.

The general principle is the one §8.4 established. A scale and a timbre belong together. Western tuning theory is the theory of scales for instruments with harmonic partials, and it is not more fundamental than the alternatives, it is more *specialized*, and its specialization is to strings and pipes.

9.7 Summary

- **An interval is a ratio**, and intervals multiply. The **cent**, 1200 to the octave, $n = 1200 \log_2 r$, makes ratios additive and is well matched to the ear, whose discrimination is 3–6 cents.
- **The chain of fifths never closes.** Twelve pure fifths exceed seven octaves by the **Pythagorean comma**, 23.5 cents, because no power of 3 is a power of 2. Four pure fifths exceed a pure major third plus two octaves by the **syntonic comma**, 21.5 cents.
- **Pythagorean tuning** gives every fifth pure and every major third 22 cents sharp. Acceptable while the third was a dissonance; untenable once it became a consonance.
- **Just intonation** gives pure triads on the tonic, subdominant, and dominant, at the cost of two sizes of whole tone, a wolf fifth at D–A, and the impossibility of modulating.
- **Quarter-comma meantone** narrows every fifth by 5.4 cents to make major thirds exactly pure, and dumps the remaining error into one unusable wolf. **Well temperaments** distribute it unevenly, making every key usable and each key different.
- **Equal temperament** makes every semitone $2^{1/12}$. Fifths are 2 cents narrow (inaudible), thirds 14 cents sharp (audible). It is a worse compromise than meantone in any one key and the only one that is the same in all of them.
- **Fixed-pitch instruments use equal temperament; flexible-pitch ones do not.** Singers and string players bend toward just intonation harmonically and away from it melodically, and negotiate with any keyboard present.
- **Pianos are tuned stretched**, because string stiffness makes their partials sharp, and an octave must match the partial rather than the arithmetic.
- **Other traditions divide the octave differently**, and where their instruments have inharmonic partials those divisions are as well founded as ours.

9.8 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple response.** Which statements are true of twelve-tone equal temperament? Consider its semitone ratio, equivalence of keys, fifth size, and major-third purity.

2. **True or false.** A chain of twelve acoustically pure fifths closes exactly at seven octaves.

3. **Drag the words.** Match Pythagorean tuning, just intonation, quarter-comma meantone, and equal temperament to their characteristic compromises.

4. **Fill in the blanks.** An octave contains _____ cents, so an equal-tempered semitone contains _____ cents.

5. **Mark the words.** Identify the performers who can continuously bend pitch toward just intonation: “A singer and a violinist can adjust each note continuously; a pianist and an organist are constrained by fixed pitches.”

9.9 Conceptual Questions

1. Explain why intervals are ratios rather than differences, and connect the answer to the cochlea.

2. Explain, in terms of prime factorization, why a chain of pure fifths can never close into a whole number of octaves, at twelve steps or at any other number.

3. Pythagorean tuning was satisfactory for medieval music and unsatisfactory for Renaissance music. Explain what changed.

4. A choir sings a sustained chord and a piano plays the same chord. Explain why the choir’s version is likely to be closer to just intonation, and why the piano’s cannot be.

5. Explain why the wolf fifth in just intonation arises from there being two sizes of whole tone.

6. Quarter-comma meantone has a fifth 5.4 cents narrow and equal temperament has one 2.0 cents narrow, yet meantone’s thirds are much better. Explain how meantone can be worse on fifths and better on thirds.

7. Explain what “key color” meant in a well temperament and why the concept largely evaporates in equal temperament.

8. A piano tuned to mathematically exact equal temperament sounds out of tune. Explain why, and say what a tuner does instead.

9.10 Problems

Exercise 9.1
(Straightforward) Convert to cents: (a) the pure fifth 3:2, (b) the pure major third 5:4, (c) the ratio 7:6.

Solution to Exercise 9.1
Using $n = 1200 \log_2 r$.
(a) $1200 \log_2(1.5) = 1200(0.58496) = 701.96$ cents.
(b) $1200 \log_2(1.25) = 1200(0.32193) = 386.31$ cents.
(c) $1200 \log_2(7/6) = 1200 \log_2(1.16667) = 1200(0.22239) = 266.87$ cents.
Therefore 702.0, 386.3, and 266.9 cents. The last lies between the equal-tempered minor third (300) and major second (200), and belongs to no Western key.

Exercise 9.2
(Straightforward) Convert to frequency ratios: (a) 700 cents, (b) 400 cents, (c) 1200 cents.

Solution to Exercise 9.2

Using $r = 2^{n/1200}$.

(a) $2^{700/1200} = 2^{0.58333} = 1.49831$.

(b) $2^{400/1200} = 2^{0.33333} = 1.25992$.

(c) $2^{1} = 2.00000$.

Therefore 1.4983, 1.2599, and 2. Comparing with the pure values 1.5 and 1.25: the tempered fifth is 0.11% narrow and the tempered third 0.79% sharp: the third is seven times worse.

Exercise 9.3

(Moderate) Calculate the Pythagorean comma from scratch. (a) Find $(3/2)^{12}$. (b) Find 2^{7}. (c) Express the ratio in cents.

Solution to Exercise 9.3

(a) $(3/2)^{12} = 3^{12}/2^{12} = 531441/4096 = 129.7463$.

(b) $2^{7} = 128$.

(c)

$$\frac{129.7463}{128} = 1.0136433, \qquad 1200 \log_2(1.0136433) = 23.46 \text{ cents}. \tag{9.15}$$

Therefore the comma is 23.5 cents, very nearly a quarter of a semitone, about four to eight times the threshold of detection, and unmistakable on sustained tones.

Exercise 9.4

(Moderate) Calculate the syntonic comma. (a) Find the Pythagorean major third $(3/2)^{4}$ reduced to one octave. (b) Compare with the pure 5:4. (c) Express the difference in cents.

Solution to Exercise 9.4

(a) $(3/2)^{4} = 81/16 = 5.0625$. Dividing by 4 to bring it into one octave: $81/64 = 1.265625$.

(b) The pure third is $5/4 = 1.25$.

(c)

$$\frac{81/64}{5/4} = \frac{81}{80} = 1.0125, \qquad 1200 \log_2(1.0125) = 21.51 \text{ cents}. \tag{9.16}$$

Therefore the syntonic comma is 21.5 cents, the ratio 81:80. It is the difference between the third you get from stacking fifths and the third the harmonic series offers, and it is what meantone temperament distributes.

Exercise 9.5

(Moderate) Take $\mathrm{C}_4 = 261.63$ Hz. Find the frequency of E_4 (a) in equal temperament, (b) in just intonation, (c) in Pythagorean tuning. (d) Express the spread in cents.

Solution to Exercise 9.5

(a) $261.63 \times 2^{4/12} = 261.63 \times 1.259921 = 329.63$ Hz.

(b) $261.63 \times 5/4 = 327.04$ Hz.

(c) $261.63 \times 81/64 = 331.12$ Hz.

(d) From lowest to highest:

$$1200 \log_2\left(\frac{331.12}{327.04}\right) = 21.5 \text{ cents}. \tag{9.17}$$

Therefore the three Es span 327.0 to 331.1 Hz, a spread of 21.5 cents: the syntonic comma, with equal temperament sitting about two-thirds of the way up.

Exercise 9.6
(Moderate) A quarter-comma meantone fifth is narrowed by a quarter of the syntonic comma. (a) By how many cents? (b) What is the meantone fifth, in cents and as a ratio? (c) Verify that four of them, less two octaves, give a pure major third.

Solution to Exercise 9.6
(a) $21.51/4 = 5.377$ cents.
(b) $701.955 - 5.377 = 696.578$ cents, which is

$$r = 2^{696.578/1200} = 1.495349. \tag{9.18}$$

(c) Four such fifths:

$$4 \times 696.578 = 2786.31 \text{ cents}. \tag{9.19}$$

Two octaves is 2400 cents, so the third is

$$2786.31 - 2400 = 386.31 \text{ cents}, \tag{9.20}$$

which is exactly $1200 \log_2(5/4)$.
Therefore the construction works exactly, and that exactness is the whole point of the quarter: it is chosen to make the third pure, not as a rough compromise.

Exercise 9.7
(Moderate) In just intonation on C, find the interval D–A. (a) Give the ratio and the cents. (b) How far is it from a pure fifth? (c) If $A_4 = 440$ Hz in this tuning, what beat rate would a listener hear between the third harmonic of D and the second harmonic of A?

Solution to Exercise 9.7
(a) D is 9/8 and A is 5/3, so

$$\frac{5/3}{9/8} = \frac{40}{27} = 1.48148, \qquad 1200 \log_2(1.48148) = 680.45 \text{ cents}. \tag{9.21}$$

(b) $701.96 - 680.45 = 21.5$ cents narrow, one syntonic comma.
(c) If A = 440 Hz then $D = 440 \times 27/40 = 297.0$ Hz. The relevant harmonics:

$$3 \times 297.0 = 891.0 \text{ Hz}, \qquad 2 \times 440 = 880.0 \text{ Hz}, \tag{9.22}$$

$$f_{\text{beat}} = 11.0 \text{ Hz}. \tag{9.23}$$

Therefore the wolf beats eleven times a second, far too fast to count and squarely in the range Chapter 8 identified as maximally rough. This is why it howls.

Exercise 9.8
(Challenging) Show that in just intonation the whole tones C–D and D–E are different sizes, and find the difference. Explain why a keyboard cannot accommodate both.

Solution to Exercise 9.8

C–D is $9/8 = 203.91$ cents. D–E is

$$\frac{5/4}{9/8} = \frac{10}{9} = 182.40 \text{ cents}. \tag{9.24}$$

The difference is

$$\frac{9/8}{10/9} = \frac{81}{80} = 21.5 \text{ cents}, \tag{9.25}$$

the syntonic comma again.

A keyboard has one key per note name, so it must assign a single pitch to D. If D is placed to make C–D a 9:8 tone, then D–E is a 10:9 tone and the fifth D–A is a comma narrow. If D is placed the other way, some other interval breaks instead.

Therefore the instrument is short of keys, not short of cleverness. Some seventeenth-century keyboards addressed this literally, with split black keys giving separate D♯ and E♭, and they were abandoned as unplayable.

Exercise 9.9

(Moderate) An equal-tempered perfect fifth is built on $\mathrm{A}_3 = 220$ Hz. (a) Find the upper frequency. (b) Find the beat rate between the third harmonic of the lower note and the second of the upper. (c) Repeat for the major third, using the fifth and fourth harmonics. (d) Comment.

Solution to Exercise 9.9

(a) $220 \times 2^{7/12} = 220 \times 1.498307 = 329.63$ Hz.

(b) $3 \times 220 = 660.00$ Hz against $2 \times 329.63 = 659.26$ Hz:

$$f_{\text{beat}} = 0.74 \text{ Hz}. \tag{9.26}$$

(c) The third: $220 \times 2^{4/12} = 277.18$ Hz. Then $5 \times 220 = 1100.0$ Hz against $4 \times 277.18 = 1108.7$ Hz:

$$f_{\text{beat}} = 8.7 \text{ Hz}. \tag{9.27}$$

(d) The tempered fifth beats less than once a second, slow enough to pass for still. The tempered third beats nearly nine times a second, which is right in the roughest range.

Therefore equal temperament's compromise is very uneven in its audible effect: the fifths are essentially free and the thirds are paid for dearly. This is exactly what the meantone temperaments refused to accept.

Exercise 9.10

(Moderate) Consider a system dividing the octave into 19 equal steps. (a) How many cents is each step? (b) Which multiple of a step best approximates a pure fifth, and with what error? (c) The pure major third?

Solution to Exercise 9.10

(a) $1200/19 = 63.16$ cents.

(b) A pure fifth is 701.96 cents:

$$\frac{701.96}{63.16} = 11.12, \tag{9.28}$$

so eleven steps, giving $11 \times 63.16 = 694.7$ cents: an error of -7.2 cents.

(c) A pure third is 386.31 cents:

$$\frac{386.31}{63.16} = 6.12, \tag{9.29}$$

so six steps, giving $6 \times 63.16 = 378.9$ cents: an error of -7.4 cents.

Therefore 19-tone equal temperament has a fifth 7 cents narrow (worse than 12-tone's 2) but a third 7 cents flat (much better than 12-tone's 14 sharp). It is, in effect, a closed version of meantone, and it was advocated on exactly those grounds in the sixteenth century and again in the twentieth.

Exercise 9.11

(Challenging) Show that no equal division of the octave can contain an exactly pure fifth.

Solution to Exercise 9.11

Suppose the octave is divided into N equal steps, so each step is the ratio $2^{1/N}$, and suppose k steps give exactly a pure fifth:

$$2^{k/N} = \frac{3}{2}. \tag{9.30}$$

Raising both sides to the power N:

$$2^k = \left(\frac{3}{2}\right)^N = \frac{3^N}{2^N}, \qquad \text{so} \qquad 2^{k+N} = 3^N. \tag{9.31}$$

The left side has only the prime factor 2; the right has only the prime factor 3. By unique factorization these can be equal only if both are 1, requiring $N = 0$, not a division of the octave at all.

Therefore no equal temperament of any size contains a pure fifth, and the same argument rules out a pure major third, since 5 is likewise not a power of 2. **Every equal temperament is a compromise**; the only choice is how good an approximation N buys, which is why 12, 19, 31, and 53 recur in the literature, they are the divisions where the approximations happen to be unusually good.

Exercise 9.12

(Moderate) A well temperament has fifths tempered by different amounts: C–G by -6 cents, G–D by -6, D–A by -4, A–E by -4, and the remaining eight fifths by whatever is needed. (a) How much of the comma has been absorbed so far? (b) How much remains to spread over the other eight? (c) Comment on the resulting key color.

Solution to Exercise 9.12

(a) $6 + 6 + 4 + 4 = 20$ cents.

(b) The Pythagorean comma is 23.5 cents, so

$$23.5 - 20 = 3.5 \text{ cents} \tag{9.32}$$

remains, spread over eight fifths, about 0.44 cents each, which is imperceptible.

(c) The heavily tempered fifths lie among the naturals, so the keys built on them, C, G, D, A, have thirds noticeably closer to pure than equal temperament provides. The remote keys, whose fifths are essentially untempered, inherit almost-Pythagorean thirds, some 20 cents sharp, and sound correspondingly tense and bright.

Therefore every key is usable and no two sound alike: the common keys are sweet and the remote ones are edgy. That is precisely what eighteenth-century writers meant by key character, and it is a real acoustic property of the instrument rather than a literary conceit.

Exercise 9.13

(Moderate) A piano's A_2 has an inharmonicity coefficient $B = 3.0 \times 10^{-4}$ and a nominal fundamental of 110 Hz. (a) Find its second partial. (b) If A_3 is tuned so that its fundamental matches that partial, what is A_3's frequency? (c) How many cents sharp of an exact octave is that?

Solution to Exercise 9.13

(a) Using $f_n = n f_1 \sqrt{1 + Bn^2}$ with $n = 2$:

$$f_2 = 2(110)\sqrt{1 + (3.0 \times 10^{-4})(4)} = 220\sqrt{1.0012} = 220.13 \text{ Hz}. \tag{9.33}$$

(b) The tuner matches A_3's fundamental to it: 220.13 Hz.

(c) An exact octave would be 220.00 Hz:

$$1200 \log_2\left(\frac{220.13}{220.00}\right) = 1.0 \text{ cent}. \tag{9.34}$$

Therefore this octave is stretched by one cent. That is small, but it *accumulates*: seven octaves tuned this way, each stretched a little and each stretching more as the strings get shorter and stiffer, produces the tens of cents of the Railsback curve at the extremes of the keyboard.

Exercise 9.14

(Moderate) A composer writes for an ensemble of a string quartet and a piano. (a) Explain the tuning conflict. (b) The quartet plays a sustained A major triad alone, then the piano enters with the same chord. Describe what a listener hears. (c) Suggest what the quartet will actually do.

Solution to Exercise 9.14

(a) The quartet can adjust its pitches continuously and will naturally tune a sustained triad toward just intonation, with the third about 14 cents flat of equal temperament. The piano is fixed in equal temperament and cannot move.

(b) Alone, the quartet's chord is still and beatless. When the piano enters, its C♯ is 14 cents sharp of the quartet's, and the two thirds beat against each other at roughly 10 Hz: a clearly audible sourness that was not there a moment before.

(c) The quartet will bend toward the piano, because it is the only party that can. In practice string players in this situation play thirds sharper than they would unaccompanied, sacrificing the purity of their own chord to avoid beating against the fixed instrument.

Therefore the compromise falls entirely on the flexible players, which is the general rule whenever fixed and flexible instruments play together, and one reason string players often describe playing with piano as harmonically constraining.

Part V

Musical Instruments

Chapter 10

String Instruments: Guitar, Violin, and Piano

Learning Objectives

By the end of this chapter, you should be able to:

- Calculate the fundamental frequency of a string from its length, tension, and linear density, and predict the effect of changing each.
- Explain why a bass string is wound rather than simply made thicker or slacker.
- Explain how the point and the manner of excitation determine which harmonics are strong, and predict the spectral effect of plucking a guitar string near the bridge rather than over the soundhole.
- Describe Helmholtz motion of a bowed string, and explain how the stick-slip mechanism sustains it.
- Explain why a string alone radiates almost no sound, and describe how the bridge and body solve the impedance-matching problem.
- Describe the principal resonances of a guitar body, including the Helmholtz air resonance, and explain what each contributes.
- Explain inharmonicity in piano strings, identify its cause in string stiffness, and connect it to stretched piano tuning.
- Explain why piano strings are struck at about one-seventh of their length, and why the piano uses multiple strings per note.
- Compare the three instrument families in terms of how energy enters the string, how long it stays, and how it leaves.

Introduction

Part V takes the machinery of the first nine chapters and applies it to instruments, one family at a time. Each chapter asks the same three questions:

1. **What vibrates?**
2. **How is energy put into it?**
3. **How does the vibration reach the air?**



Figure 10.1: The string is only the beginning: the bridge and body are the machinery that turns a small string motion into radiated sound. The Metropolitan Museum of Art, CC0.

For string instruments, the first answer is already known. Chapter 3 derived the modes of a string fixed at both ends, and found $f_n = (n/2L)\sqrt{T/\mu}$: a complete harmonic series, so a string has a definite pitch and a good one.

The other two questions are where the instruments differ, and where the interesting physics lies.

The **second** question, how energy gets in, turns out to control the timbre almost entirely. Chapter 4 showed that a pluck decomposes into modes, and that the pluck point determines which. Plucking, bowing, and striking excite quite different spectra from the same string, and within each, exactly where you do it matters as much as how hard.

The **third** question is a problem rather than a mechanism. A vibrating string, on its own, is almost silent. Chapter 2 gave the reason: the impedance mismatch between a thin dense string and thin light air is enormous, and almost no energy crosses. Everything else about a string instrument, the bridge, the body, the soundboard, the shape, the wood, exists to solve that problem, and the solutions are what make a guitar sound different from a violin.

10.1 The Vibrating String

Frequency, Length, Tension, and Mass

The governing formula, from Chapter 3:

$$f_n = \frac{n}{2L}\sqrt{\frac{T}{\mu}}. \tag{10.1}$$

Three levers, and each has a characteristic use.

Length is the player's lever. A finger on a fingerboard or a fret shortens the string, and pitch goes as $1/L$: the only one of the three that is a simple inverse, and so the only one suited to rapid, accurate, continuous control.

Tension is the tuner's lever. Pitch goes as $\sqrt{T}$, so an octave costs four times the tension. Strings already run close to their breaking stress, so this lever has very little travel, which is exactly what makes it good for fine tuning.

Mass per unit length is the maker's lever. Pitch goes as $1/\sqrt{\mu}$, so a string four times as heavy sounds an octave lower at the same length and tension.

Worked example: why bass strings are wound

A guitar's high E string is 0.25 mm in diameter. Its low E is two octaves lower, at the same length and a similar tension. How thick would a plain steel string need to be, and what goes wrong?

Two octaves is a factor of four in frequency, so μ must rise by a factor of 16. For a solid cylinder $\mu \propto d^2$, so the diameter must rise by $\sqrt{16} = 4$:

$$d = 4 \times 0.25 \text{ mm} = 1.0 \text{ mm}. \tag{10.2}$$

That is a steel rod a millimeter thick. Two things go wrong with it.

It is far too stiff to bend. Chapter 5 showed that stiffness makes the partials inharmonic, and the inharmonicity coefficient rises steeply with diameter: a 1 mm plain string would be audibly and unpleasantly out of tune with itself.

It will not bend over the bridge or turn round a tuning peg without taking a permanent set.

The solution is a **wound** string: a thin, flexible core carrying the tension, with a heavy wire wrapped loosely around it adding mass. The winding contributes μ without contributing bending stiffness, because the turns are free to move relative to one another. A wound low E achieves the sixteenfold mass with a core no thicker than about 0.4 mm.

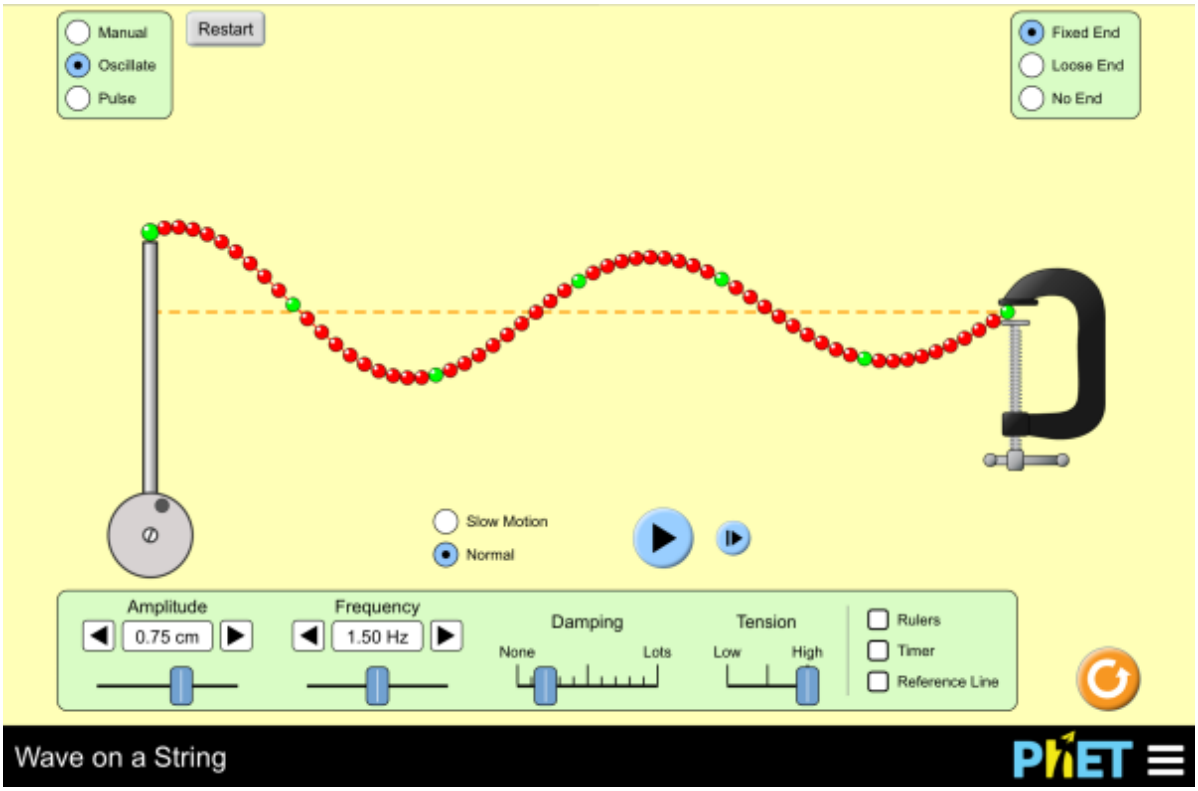


Figure 10.2: Pluck a string and watch the pulse reflect. Two things are worth doing deliberately: switch the far end between fixed and loose and confirm the inversion of Chapter 3, and then set the damping to zero and drive the string to see the standing-wave modes appear one at a time. Interactive simulation: Wave on a String

The Mode Shapes

The modes are $\sin(n\pi x/L)$, and Chapter 4 established the rule that matters most for what follows:

> **Exciting the string at a point excites only those modes that move at that point.**

Since the nth mode has nodes at $x = L/n, 2L/n, \ldots$, touching or driving at L/n affects mode n not at all.

Real Strings: Stiffness and Inharmonicity

An ideal string resists only stretching. A real one also resists **bending**, and that extra restoring force stiffens the high modes more than the low ones, because they are more sharply curved. Chapter 5 gave the result:

$$f_n = n f_1 \sqrt{1 + B n^2}. \tag{10.3}$$

The coefficient B grows with the fourth power of the string's diameter and falls with the square of its length, so **short thick strings are the worst**. That combination describes the extremes of a piano, and §10.6 follows the consequence.

10.2 Setting the String in Motion

Plucking: Where You Pluck Is What You Hear

A pluck displaces the string into a triangle and releases it. Chapter 4 derived the mode amplitudes:

$$c_n \propto \frac{1}{n^2} \sin\!\left(\frac{n\pi a}{L}\right), \tag{10.4}$$

with a the distance of the pluck from one end.

Two factors, doing two different jobs. The $1/n^2$ makes high harmonics weak whatever you do, this is why a plucked string is not harsh. The $\sin(n\pi a/L)$ is the interesting one: it vanishes whenever $a = L/n$, so **plucking at L/n removes the nth harmonic and all its multiples**, and it also tilts the whole spectrum according to where a is.

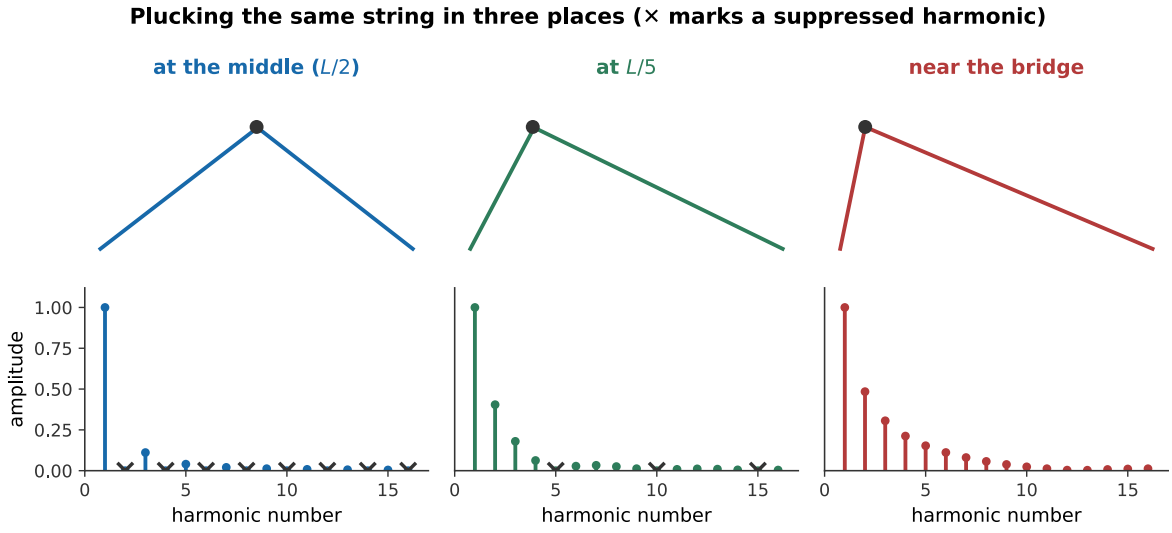


Figure 10.3: Three pluck positions on one string. At the middle, every even harmonic vanishes and the tone is hollow and mellow. At $L/5$, only the 5th and 10th are lost. Near the bridge, no low harmonic is suppressed and the high ones are relatively strong, which is the bright, thin, nasal sound.

The guitar vocabulary follows directly: *sul tasto* or "over the fingerboard" means plucking near the middle for a mellow sound; *sul ponticello* or "at the bridge" means plucking near the end for a bright one.

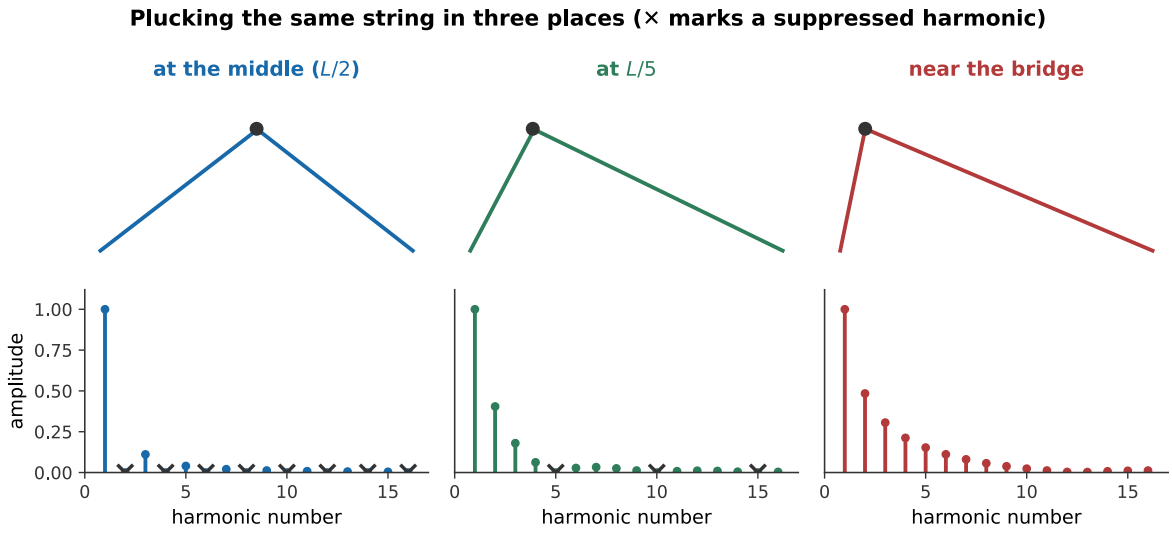


Figure 10.4: The same string, the same pitch, the same loudness, three right-hand positions. A guitarist's right hand moves as much as the left for that reason, and it is Chapter 4's mode-decomposition made audible. *What you hear:*The same string plucked three times. The first is soft and hollow; the second full and balanced; the third thin, bright and nasal, like a harpsichord. Listen: At the middle, At L/5, Near the bridge

Bowing and Helmholtz Motion

Bowing is not plucking repeated. It is a genuinely different mechanism, and the first person to work out what actually happens was Helmholtz, using a vibration microscope.

The intuition that the bow drags the string smoothly along is wrong. What happens instead is **stick–slip**:

1. The rosined bow hair grips the string by static friction and carries it along at the bow's speed.
2. The string's restoring force grows until it exceeds what static friction can hold.
3. The string breaks free and slips back rapidly, in the opposite direction, past the bow.
4. The bow catches it again, and the cycle repeats.

The remarkable part is the *shape* the string takes while this happens.

Two consequences follow, and both matter.

The pitch is set by the string, not by the bow. One slip occurs per round trip of the kink, and that round trip takes exactly one period of the fundamental. Bow speed and pressure change the loudness and the timbre; they do not change the pitch.

The spectrum is a sawtooth. The force the string applies to the bridge is proportional to the string's slope there, and the kink passing the bridge makes that slope jump abruptly once per cycle. A sawtooth contains every harmonic, falling as $1/n$, much richer than a pluck's $1/n^2$, so a bowed note is brighter and more penetrating than a plucked one.

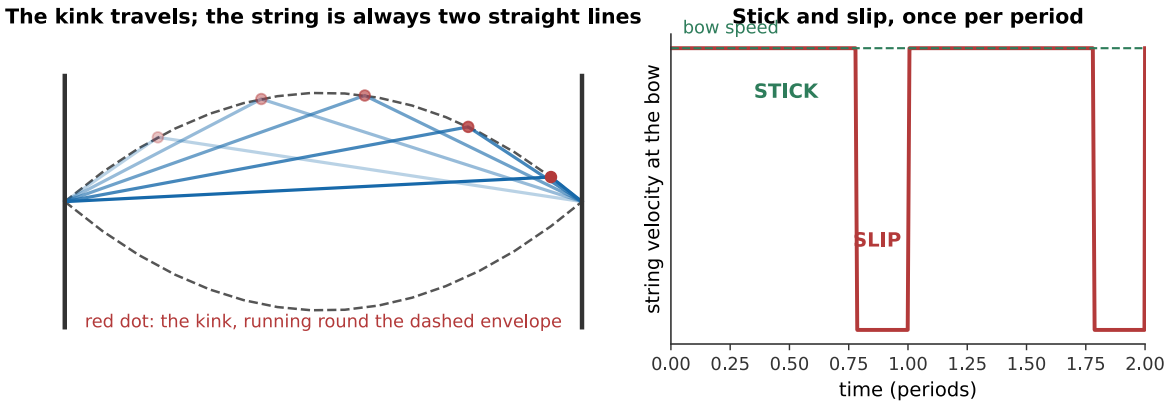


Figure 10.5: Helmholtz motion. At every instant the string is exactly two straight lines meeting at a sharp kink, and the kink travels around a parabolic envelope once per period. The right panel shows the consequence at the bow: the string moves with the bow for most of the cycle and slips back quickly once, one slip per period, which is what sets the pitch.

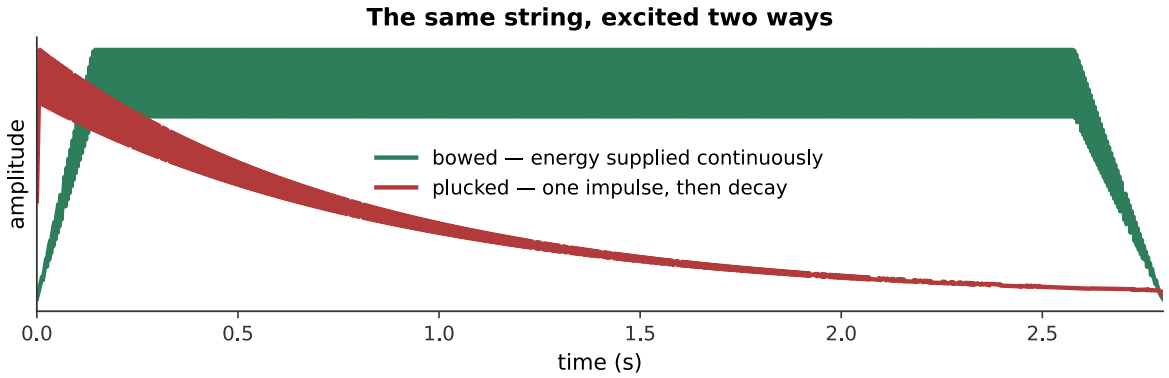


Figure 10.6: The same string, excited two ways. Note that two things differ, and they are independent: the envelope (sustained against decaying) and the spectrum ($1/n$ against $1/n^2$). A synthesizer that gets one right and the other wrong is instantly recognizable as wrong. *What you hear:*The same pitch twice. The first grows in, holds steady and is bright and sustained; the second starts immediately and dies away, and is mellower. Listen: Bowed, Plucked

Note

Helmholtz motion only establishes itself within a certain range of bow force and bow position, mapped out by John Schelleng in 1973. Too little force and the string never sticks: the result is a whistling, unfocused sound called *surface sound*. Too much and the string sticks too long, producing a raucous crunch. The window narrows sharply as the bow moves toward the bridge, which is precisely why *sul ponticello* is difficult to control and why beginners sound scratchy.

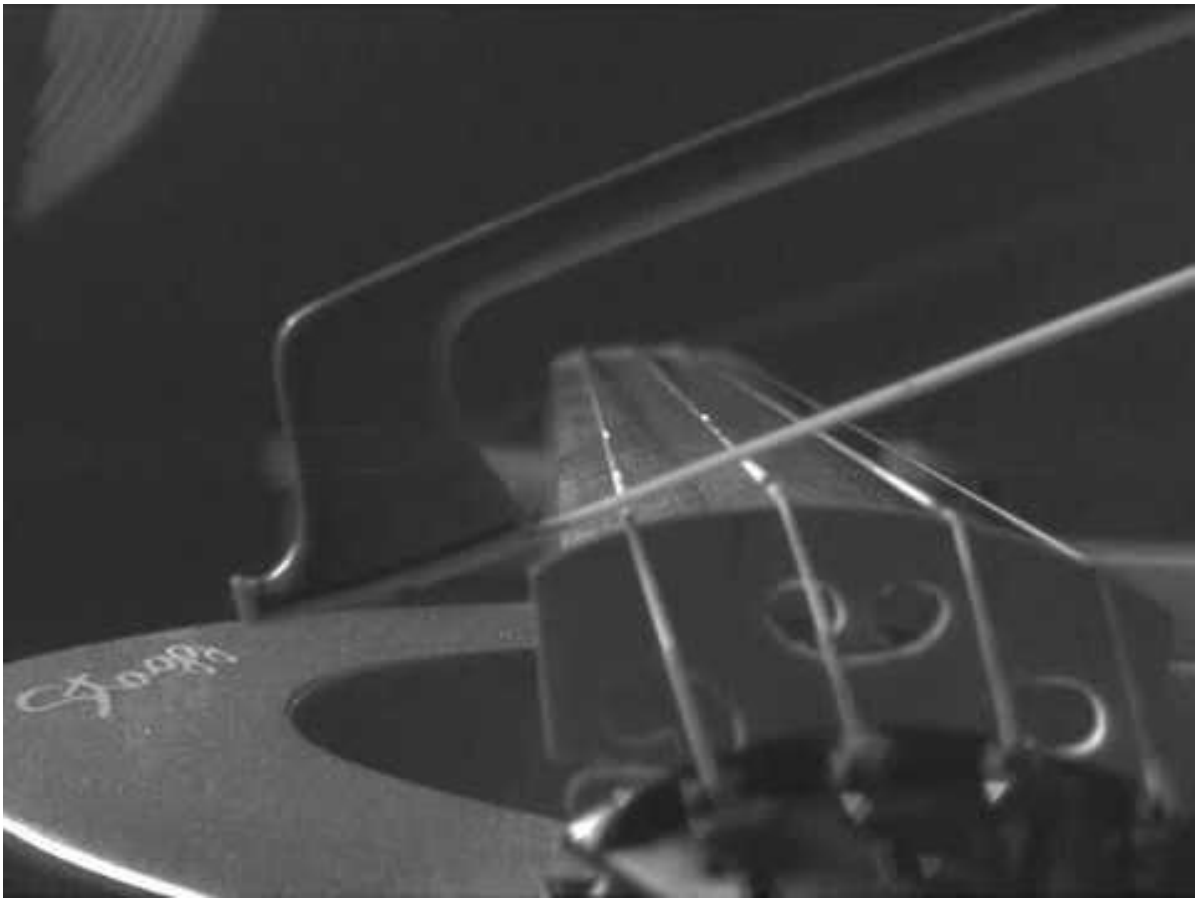


Figure 10.7: High-speed footage reveals Helmholtz motion: a sharp corner travels back and forth along the bowed string as sticking alternates with slipping. The motion that produces the sawtooth-like waveform is difficult to see at normal speed. Watch on YouTube: Bowed Violin String in Slow Motion

Striking: The Hammer and the Contact Time

A piano hammer is the third mechanism. It differs from a pluck in two ways that both matter.

The contact time is finite. A hammer stays on the string for a millisecond or two, long enough to span a significant fraction of the period of the high harmonics. Any harmonic whose half-period is shorter than the contact time is suppressed, because the hammer is still there when it tries to come back. The hammer therefore acts as a low-pass filter, and a soft hammer, which stays in contact longer, produces a mellower note than a hard one.

The hammer is nonlinear. Felt compresses stiffer the harder it is squeezed, so a loud blow has a shorter contact time than a quiet one. The consequence is that **a piano played louder is also brighter**, not merely louder: a coupling between dynamics and timbre that is a large part of what the instrument is expressive with.

10.3 From String to Air

The Impedance Problem Restated

Now the third question. A string moving in air pushes very little air: it is thin, so it cuts through rather than driving, and the air offers almost no resistance to being pushed. In the language of Chapter 2, the string's impedance is far higher than the air's, so almost all the energy reflects back into the string.

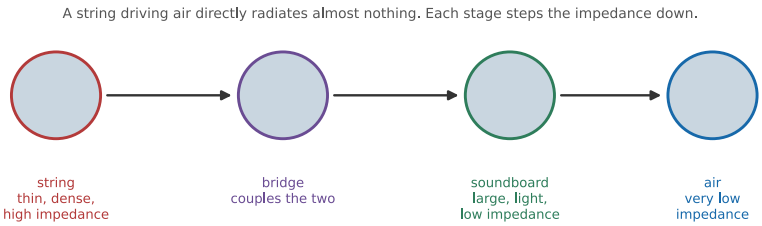


Figure 10.8: The chain. No single step bridges the gap between a steel string and open air; the instrument steps down the impedance in stages, and each stage is a piece of the instrument a maker can adjust.

A useful check on how bad the mismatch is: an electric guitar has a solid body and therefore an almost perfect mismatch. Unplugged, it is barely audible, and it sustains for a very long time, because the energy has nowhere to go. That is Chapter 4's trade-off in its purest form. It sidesteps the mismatch entirely rather than solving it, reading the string's motion magnetically and rebuilding the sound electrically; how that pickup works is Chapter 15's subject.

The Bridge

The **bridge** is the coupling element, and it does two jobs.

It **transmits** the string's transverse force to the top plate. The string's tension pulls down over the bridge, and as the string vibrates, that downward force varies, driving the plate.

It **filters**. A bridge has mass and stiffness of its own, so it has resonances, and it passes some frequencies better than others. A violin bridge has a strong resonance around 3 kHz, the **bridge hill**, which lands squarely in the region of greatest hearing sensitivity (Chapter 7) and contributes a good deal of the violin's carrying power.

Bridge design is therefore a filter-design problem, and violin makers adjust it by carving: thinning the bridge's waist lowers its resonance, and cutting away the "kidneys" changes its coupling.

Soundboards, Bodies, and Radiation

The **soundboard** is the actual radiator. Its job is to present a large, light surface to the air.

- **Large**, because a small source radiates poorly at low frequencies: the wavelengths involved are meters and a small body cannot get a grip on them (Chapter 2).

- **Light**, because a heavy plate is hard for the string to drive.
- **Stiff**, because a floppy plate will not move as a unit.

Light and stiff are contradictory in most materials, and this is why instrument soundboards are made of spruce and have been for five hundred years. Spruce has an exceptionally high stiffness-to-density ratio along the grain, and Chapter 2 noted that sound travels along the grain some four times faster than across it. Soundboards are cut with the grain running the length of the instrument for exactly that reason.

10.4 The Guitar

Scale Length, Frets, and Intonation

A guitar's frets are placed so that stopping at fret n shortens the string by a factor of $2^{n/12}$: each fret is a semitone, and the twelfth fret sits at exactly half the string length.

The scheme has a systematic error. Pressing a string to a fret **stretches** it slightly, raising its tension and therefore its pitch. The effect is worse higher up the neck, where the string must be pushed further, and worse on thick strings.

The remedy is **compensation**: the bridge saddle is set slightly *further* from the nut than the theoretical scale length, and by a different amount for each string, so that the stretching error cancels. Look at an electric guitar's bridge and the six saddles are visibly staggered, that stagger is this correction.

Body Resonances and the Helmholtz Air Mode

A guitar body's response is not flat.

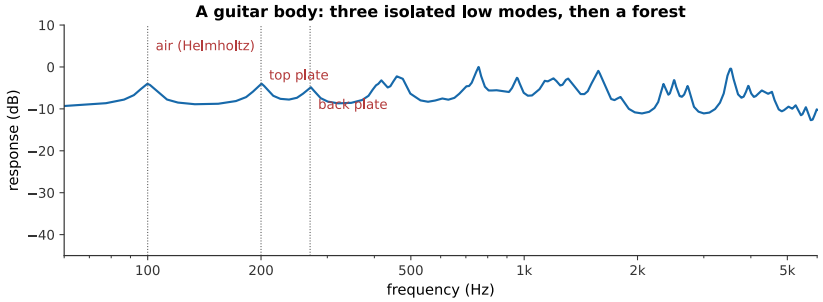


Figure 10.9: A guitar body's frequency response. Below about 300 Hz there are a few isolated, individually identifiable modes. Above it, the modes become so numerous and closely spaced that they merge into a lumpy continuum. Makers can tune the low modes deliberately; the forest above can only be influenced statistically.

The lowest peak is the **Helmholtz air resonance** of Chapter 4: the body cavity and soundhole acting as a mass of air bouncing on a spring of air, typically around 100 Hz on a steel-string guitar. It is deliberately placed near the bottom of the instrument's range, where the top plate alone radiates poorly, and it is why a guitar has a soundhole rather than merely a hollow body.

Above it come the plate modes: the top plate flexing, the back plate flexing, and the two coupled through the air and through the sides.

Why Two Guitars Sound Different

Everything in this chapter so far applies equally to any guitar. What distinguishes one instrument from another is almost entirely the **filter**: the placement, strength, and Q of the body's resonances.

The bracing pattern under the top plate is therefore the central object of guitar design, and makers argue about it. The braces set the top's stiffness distribution, which sets its mode shapes and frequencies, which sets the filter. Thin a brace by half a millimeter and the filter changes.

It is also why the argument is hard to settle. Chapter 4's trade-off is unavoidable, bracing for volume costs sustain and evenness, and different makers and players weigh those differently.

10.5 The Violin Family

Bowing in Practice

A violinist controls three things continuously: bow **speed**, bow **force**, and the bow's **distance from the bridge**. Together they determine where the note sits in Schelleng's diagram, and therefore both its loudness and its timbre.

The rule of thumb is that these three must be traded against one another. Playing close to the bridge permits a very loud, bright tone but demands high force and allows little error. Playing over the fingerboard is forgiving and produces a flute-like sound that cannot be made loud.

The Bridge, the Soundpost, and the Bass Bar

Inside a violin are two pieces of wood that do not touch the strings and are essential.

The **soundpost** is a spruce dowel wedged between the top and back plates under the treble foot of the bridge. It does two things at once. Structurally, it resists the downward force of the strings, some 90 N pressing on the bridge. Acoustically, it pins one foot of the bridge so that the bridge **rocks** about it rather than moving up and down as a whole, which is what converts the string's motion into an efficient drive of the top plate. Moving a soundpost by half a millimeter audibly changes an instrument, and it is the first thing a luthier adjusts.

The **bass bar** is a spruce strip glued under the top on the bass side. It stiffens the plate lengthwise, spreads the bridge's drive along the instrument, and lowers the frequency of the important low modes.

The Bridge Hill, Air Modes, and the Wolf Note

Three named features of the violin's response are worth knowing.

The **air mode** near 280 Hz is the Helmholtz resonance of the body through the f-holes, and it supports the bottom of the instrument's range, close to the open G string.

The **bridge hill** near 3 kHz is the bridge's own resonance, and it is the violin's projection. It coincides with the ear's most sensitive region and with the singer's formant of Chapter 13, several unrelated traditions have independently converged on putting energy at 3 kHz, because that is where the ear is listening.

The **wolf note** is a defect rather than a feature. If a strongly radiating body resonance lies at the same frequency as a note the player wants, the body absorbs energy from the string so efficiently that the string cannot maintain steady Helmholtz motion. The two alternately



Figure 10.10: The instrument's visible parts are a map of its acoustic compromises: flexible strings, a light bridge, thin plates, and a resonant cavity. Smithsonian Institution, CC0.

grab the energy, and the result is a stuttering, howling note, usually around E or F on a cello. The standard remedy is a **wolf eliminator**, a small mass clamped to the string beyond the bridge, which splits the offending resonance into two weaker ones on either side.

10.6 The Piano



Figure 10.11: Instrument design is a sequence of compromises between stiffness, mass, damping, and radiation. This piano, built by Bartolomeo Cristofori around 1720, is the oldest surviving instrument of his invention. The Metropolitan Museum of Art, CC0.

A String Under Great Tension

A modern piano is an extreme instrument. Its strings carry a total tension of around 200 kN, some twenty tonnes, so it needs a cast-iron frame, and the modern piano could not exist before the industrial techniques to cast one.

The range is seven and a quarter octaves, a factor of 150 in frequency, and the formula $f = (1/2L)\sqrt{T/\mu}$ cannot cover that with length alone: the lowest string would have to be over ten meters. So all three levers are used at once: the bass strings are longer, heavier (wound), and at somewhat lower tension.

Hammers, Strike Point, and Unisons

Piano hammers strike at about 1/7 to 1/8 of the string's length, and the choice is deliberate. §10.2's rule says this suppresses the 7th (or 8th) harmonic and its multiples, and the 7th harmonic is 31 cents flat of any equal-tempered note (Chapter 9), so it clashes with every chord the instrument plays. Striking at $L/7$ removes it.

Most notes have **two or three strings** tuned to the same pitch. The reason is not simply loudness: three strings give $10\log_{10} 3 = 4.8$ dB, which is a small return for tripling the hardware. The real reason is **sustain**. A single string coupled to a soundboard efficient enough to be

loud dies very quickly. Three strings, very slightly detuned and coupled through the bridge, exchange energy back and forth, and the note decays in two stages, a fast initial fall followed by a long, quiet "aftersound". That two-stage decay is one of the most recognizable features of a piano, and the *una corda* pedal, which shifts the action so the hammer strikes fewer strings, changes it deliberately.

Inharmonicity and Stretched Tuning

Finally, the consequence of §10.1's stiffness.

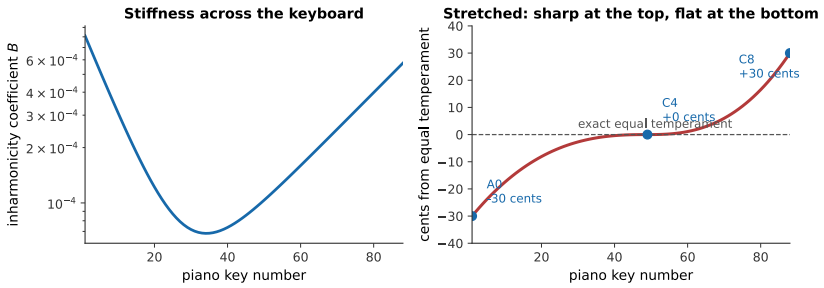


Figure 10.12: Left: string stiffness across the keyboard, worst at the extremes, the bass strings because they are thick, the treble because they are short. Right: the **Railsback curve**, the stretched tuning that results. The curve shown is representative of measured tunings rather than computed from the left panel, since how much a tuner stretches depends on which partials they choose to match.

The mechanism: a tuner setting an octave listens for beats between the lower note's **second partial** and the upper note's fundamental. Because of inharmonicity that second partial is *sharp* of twice the fundamental, so a beatless octave is wider than 2:1. Repeat up and down the keyboard and the errors accumulate.

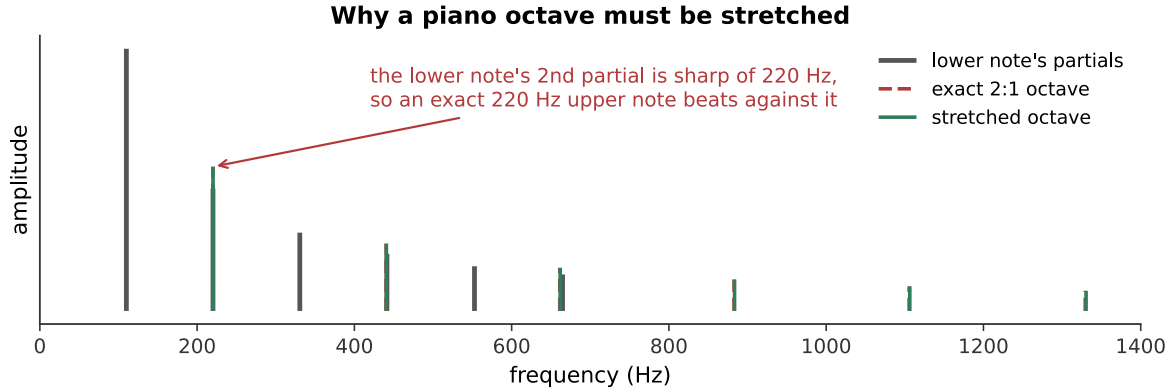


Figure 10.13: The mathematically exact octave is the one that beats. This is the reverse of what a reader expects, and it is the whole justification for the Railsback curve: on an instrument with inharmonic partials, "in tune" means matching the partials that are actually there rather than the ratio that ought to be. *What you hear:*Two octaves on a simulated piano. The first, tuned to an exact 2:1, has a clear slow beating in it. The second, tuned slightly wide, is still. Listen: Exact 2:1 octave, Stretched octave

A well-tuned concert grand is therefore about 30 cents sharp at the top and 30 cents flat at the bottom: a spread of more than half a semitone across the instrument. Nobody hears

it as out of tune. Played against an exactly equal-tempered synthesizer, the piano is the one that sounds right.

10.7 Summary

- **A string's modes are harmonic**, $f_n = (n/2L)\sqrt{T/\mu}$. Length is the player's lever $(1/L)$, tension the tuner's ($\sqrt{T}$, so an octave costs four times), mass the maker's $(1/\sqrt{\mu})$.
- **Bass strings are wound** because a plain string heavy enough would be too stiff, inharmonic and unbendable. Winding adds mass without adding bending stiffness.
- **Pluck position sets the spectrum**: $c_n \propto n^{-2}\sin(n\pi a/L)$, so plucking at L/n removes the nth harmonic and its multiples, and plucking near the bridge tilts the spectrum toward the high harmonics.
- **Bowing produces Helmholtz motion**: stick–slip, with the string always two straight lines meeting at a kink that circulates once per period. The pitch is set by the string; the force at the bridge is a sawtooth, giving a $1/n$ spectrum.
- **A hammer's finite contact time low-passes the spectrum**, and felt's nonlinearity shortens contact at higher force, so a piano played louder is also brighter.
- **The impedance mismatch between string and air is the central problem.** Bridge and soundboard step it down; an electric guitar, which does not, is nearly silent and sustains for a very long time.
- **The guitar's lowest body resonance is a Helmholtz air mode** near 100 Hz. Above about 300 Hz the body's modes merge into a forest, and what distinguishes two instruments is the shape of that filter.
- **The violin's soundpost makes the bridge rock rather than translate**, which is what couples the string to the plate efficiently. The **bridge hill** near 3 kHz gives projection, in the ear's most sensitive region.
- **Piano hammers strike at about** $L/7$ to suppress the seventh harmonic, which is 31 cents flat of any key. Multiple strings per note exist mainly for the two-stage decay they produce, not for loudness.
- **Piano tuning is stretched** because string stiffness makes partials sharp: a beatless octave is wider than 2:1, and the accumulated result is the Railsback curve, about ±30 cents at the extremes.

10.8 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple choice.** With length and linear density fixed, by what factor must tension change to raise a string by one octave: $\sqrt{2}$, 2, or 4?

2. **True or false.** An unplugged solid-body electric guitar is quiet and sustains well because it lacks an efficient soundboard coupling the strings to air.

3. **Drag the words.** Complete the effects of shorter length, greater tension, greater linear density, and winding on a string.

4. **Fill in the blanks.** Plucking at one-fifth of a string's length suppresses the ____ harmonic and its ____.

5. **Mark the words.** Identify the two components that chiefly couple an acoustic string to air: “The string drives the bridge, which drives the soundboard; the large moving surface then radiates efficiently.”

10.9 Conceptual Questions

1. Explain why instruments change pitch mainly by changing length rather than tension, using the exponents in $f = (1/2L)\sqrt{T/\mu}$.

2. Explain why a bass guitar string is wound rather than simply made thicker, giving two distinct reasons.

3. A guitarist plucks near the bridge and the tone becomes bright and thin. Explain this in terms of mode amplitudes rather than by saying “it excites more harmonics”.

4. Explain why the pitch of a bowed note does not depend on bow speed, given that bow speed clearly affects loudness.

5. Explain why a bowed note is brighter than a plucked note on the same string, referring to the two spectra.

6. A piano played loudly sounds brighter as well as louder. Explain the mechanism, and say why this would not happen with a rigid hammer.

7. An unplugged electric guitar is quiet and sustains for a long time. Explain both facts with one argument.

8. Explain why the octave that beats on a piano is the mathematically exact one, and what a tuner does instead.

10.10 Problems

Exercise 10.1
(Straightforward) A guitar string is 64.8 cm long with $\mu = 4.3 \times 10^{-3}$ kg/m, tuned to 110 Hz. (a) What is its tension? (b) What tension would tune it to 146.8 Hz?

Solution to Exercise 10.1
(a) From $f_1 = (1/2L)\sqrt{T/\mu}$, so $T = \mu(2Lf_1)^2$:

$$T = (4.3 \times 10^{-3})\,[2(0.648)(110)]^2 = (4.3 \times 10^{-3})(142.6)^2 = 87.4 \text{ N}. \tag{10.5}$$

(b) Tension goes as f^2:

$$T' = 87.4\left(\frac{146.8}{110}\right)^2 = 87.4(1.781) = 156 \text{ N}. \tag{10.6}$$

Therefore about 87 N and 156 N. Note that a musical fourth, five semitones, needs nearly double the tension, so this is not how instruments are played.

Exercise 10.2
(Straightforward) A string is plucked at 1/6 of its length. (a) Which harmonics are absent? (b) If the fundamental is 147 Hz, what are the frequencies of the two lowest absent harmonics? (c) Where would you pluck to remove the 4th harmonic?

Solution to Exercise 10.2
(a) The 6th and its multiples: 6, 12, 18, …
(b) $6 \times 147 = 882$ Hz and $12 \times 147 = 1764$ Hz.
(c) At $L/4$, which is a node of the 4th mode.
Therefore the pluck point acts as a comb, and the guitarist choosing a right-hand position is choosing which teeth of the comb to apply.

Exercise 10.3
(Moderate) A guitar's high E string is 0.23 mm in diameter and sounds 330 Hz. A plain steel string of the same length and tension is wanted at 82.4 Hz. (a) By what factor must μ increase? (b) What diameter would that need? (c) Comment.

Solution to Exercise 10.3
(a) $f \propto 1/\sqrt{\mu}$, and $330/82.4 = 4.0$, so

$$\mu' = 4.0^2\,\mu = 16\,\mu. \tag{10.7}$$

(b) For a solid cylinder $\mu \propto d^2$:

$$d' = \sqrt{16}\,(0.23\ \text{mm}) = 0.92\ \text{mm}. \tag{10.8}$$

(c) Nearly a millimeter of solid steel. The inharmonicity coefficient goes as d^4, so it would rise by a factor of $16^2 = 256$, making the string audibly inharmonic, and it would be far too stiff to bend over the bridge. Hence winding.

Exercise 10.4
(Straightforward) A violin's A string sounds 440 Hz, is 32.5 cm long, and is bowed. (a) How many times per second does the kink complete a round trip? (b) How long does one round trip take? (c) What is the wave speed on the string?

Solution to Exercise 10.4
(a) Once per period, so 440 times per second.
(b) $T = 1/440 = 2.27 \times 10^{-3}$ s.
(c) The kink travels the length twice in one period:

$$v = \frac{2L}{T} = 2Lf = 2(0.325)(440) = 286\ \text{m/s}. \tag{10.9}$$

Therefore the kink makes 440 round trips a second at 286 m/s, which is, as it must be, the same wave speed the standing-wave formula gives.

Exercise 10.5
(Moderate) A piano hammer is in contact with a string for 1.3 ms. (a) Which harmonics of a 262 Hz note have half-periods shorter than the contact time? (b) What does this do to the spectrum? (c) How does a harder hammer change the answer?

Solution to Exercise 10.5

(a) The nth harmonic has period $1/(262n)$ and half-period $1/(524n)$. Setting this below 1.3 ms:

$$\frac{1}{524n} < 1.3 \times 10^{-3} \;\Rightarrow\; n > \frac{1}{524 \times 1.3 \times 10^{-3}} = 1.47. \tag{10.10}$$

So harmonics from the 2nd upward are affected, increasingly so as n rises.

(b) The hammer is still in contact while those harmonics try to reverse, so it damps them. The effect is a low-pass filter whose cutoff is around $1/(2 \times 1.3 \text{ ms}) = 385$ Hz and which rolls off progressively above it.

(c) A harder hammer compresses less and leaves sooner, raising the cutoff and letting more high harmonics through: a brighter note. Voicing a piano is therefore done by pricking the hammer felt with needles to soften it, or by hardening it with lacquer.

Exercise 10.6

(Moderate) A piano string sounds 110 Hz with $B = 3.5 \times 10^{-4}$. (a) Find the frequency of its 2nd partial. (b) To what frequency must the octave above be tuned for a beatless octave? (c) How many cents wide of 2:1 is that?

Solution to Exercise 10.6

(a) Using $f_n = nf_1\sqrt{1 + Bn^2}$ with $n = 2$:

$$f_2 = 220\sqrt{1 + (3.5 \times 10^{-4})(4)} = 220\sqrt{1.0014} = 220.154 \text{ Hz}. \tag{10.11}$$

(b) The upper note's fundamental must match it: 220.154 Hz.

(c)

$$1200 \log_2\left(\frac{220.154}{220}\right) = 1.2 \text{ cents}. \tag{10.12}$$

Therefore this octave must be stretched by about 1.2 cents. A single octave's stretch is small; seven of them, each a little larger than the last as the strings get stiffer, is what produces the Railsback curve.

Exercise 10.7

(Moderate) A guitar body has a Helmholtz air resonance at 98 Hz with $Q = 6$, and the guitar's lowest note is 82.4 Hz. (a) Find the bandwidth. (b) Is the lowest note inside it? (c) Comment on the design choice.

Solution to Exercise 10.7

(a) $\Delta f = f_0/Q = 98/6 = 16.3$ Hz, so the half-power band runs from about 89.8 to 106.2 Hz.

(b) No: 82.4 Hz is about 7 Hz below the lower half-power point.

(c) The resonance sits just *above* the lowest note rather than on it, which is the usual arrangement. Placing it exactly on the lowest note would make that one note boom and the ones above it comparatively weak. Placing it a little higher lets its skirt support the bottom note while the peak reinforces the notes just above, giving a more even bass across several semitones: the low-Q design philosophy of Chapter 4 applied deliberately.

Exercise 10.8
(Challenging) A cello has a wolf note at 174.6 Hz (F_3). (a) Explain the mechanism. (b) A wolf eliminator of mass 8 g is fitted to the string beyond the bridge. Explain qualitatively why adding a mass helps.

Solution to Exercise 10.8
(a) A strong, well-radiating body resonance coincides with F_3. The body absorbs energy from the string so efficiently at that frequency that the string cannot sustain steady Helmholtz motion: the string builds up, dumps energy into the body, loses its stick–slip regime, re-establishes it, and repeats. The result is a stuttering warble at a few hertz.

(b) The eliminator adds a small resonant system coupled to the offending one. Two coupled resonators, as Chapter 4 showed, have two normal modes with frequencies split either side of the original; so the single strong peak at 174.6 Hz becomes two weaker peaks, neither of which absorbs enough energy to break the Helmholtz regime.

Therefore the fix is not damping the string but splitting the body's resonance, and the mass must be tuned: too light or too heavy and the split is in the wrong place.

Exercise 10.9
(Moderate) A piano note has three strings, each producing 70 dB alone. (a) What level do all three give? (b) Why is this a poor argument for having three strings? (c) What is the real reason?

Solution to Exercise 10.9
(a) Three incoherent sources add $10\log_{10}3 = 4.8$ dB:

$$L = 70 + 4.8 = 74.8 \text{ dB}. \tag{10.13}$$

(b) Under 5 dB is barely more than half a doubling of loudness (Chapter 7), for three times the strings, three times the tension on the frame, and three times the tuning work. As a loudness argument it is feeble.

(c) **Sustain.** A single string coupled tightly enough to the soundboard to be loud would decay very fast. Three slightly detuned strings coupled through the bridge exchange energy with one another, producing a fast initial decay followed by a much longer aftersound. That two-stage decay is what a piano sounds like, and it cannot be had from one string.

Exercise 10.10
(Challenging) Show that plucking a string at $a = L/2$ removes every even harmonic, and explain why the resulting tone is the mellowest available.

Solution to Exercise 10.10
The mode amplitude is $c_n \propto n^{-2}\sin(n\pi a/L)$. With $a = L/2$:

$$\sin\!\left(\frac{n\pi}{2}\right) = \begin{cases} \pm 1 & n \text{ odd} \\ 0 & n \text{ even}. \end{cases} \tag{10.14}$$

So every even harmonic vanishes.

The tone is mellowest for two reasons acting together. First, half the partials are gone outright, which removes the octave, the second octave, and so on: the same hollowing that distinguishes a stopped pipe in Chapter 3. Second, the surviving odd harmonics still carry the $1/n^2$ factor, so they are weak: the 3rd is at 1/9 of the fundamental, the 5th at 1/25.

Therefore the midpoint pluck gives the fewest and weakest partials of any pluck position, which is what "mellow" means spectrally.

Exercise 10.11
(Challenging) A violin string is bowed at 1/11 of its length from the bridge. (a) Which harmonic does this most affect, and how? (b) Helmholtz motion still produces a full harmonic series. Reconcile these two statements.

Solution to Exercise 10.11
(a) The bow sits at a node of the 11th mode, so it cannot drive that mode.

(b) The reconciliation is that bowing is a *sustained, nonlinear* process, not a single impulse. The stick–slip cycle is locked to the round trip of the kink, and that round trip forces the whole motion to be periodic at f_1. Fourier's theorem (Chapter 5) then guarantees that a periodic waveform has a complete harmonic series: the 11th harmonic is present because periodicity requires it, even though the bow is not driving it directly.

Therefore the bow position affects the *balance* of the harmonics, and quite strongly, there is an audible weakening around the 11th, but it cannot delete one, as a pluck can. This is a real difference between impulsive and sustained excitation, and it is why the "pluck at a node" rule does not transfer to bowing unchanged.

Exercise 10.12
(Moderate) A grand piano's strings carry a total tension of 200 kN. (a) Express this as a mass under gravity. (b) The frame is cast iron with a tensile strength of about 200 MPa. What cross-sectional area is needed for a safety factor of 4?

Solution to Exercise 10.12
(a) From $F = mg$:

$$m = \frac{200 \times 10^3 \text{ N}}{9.81 \text{ m/s}^2} = 2.0 \times 10^4 \text{ kg} = 20 \text{ tonnes}. \tag{10.15}$$

(b) A safety factor of 4 means designing to $200/4 = 50$ MPa:

$$A = \frac{F}{\sigma} = \frac{200 \times 10^3 \text{ N}}{50 \times 10^6 \text{ Pa}} = 4.0 \times 10^{-3} \text{ m}^2 = 40 \text{ cm}^2. \tag{10.16}$$

Therefore about 40 cm^2 of iron in cross-section: a bar roughly 6 cm square, or the equivalent spread across the frame's several members. A grand piano weighs several hundred kilograms for that reason, and the instrument in its modern form had to wait for industrial iron casting.

Exercise 10.13
(Moderate) Two guitars are identical except that one has a body Q of 4 and the other Q of 9 at its main air resonance. (a) Compare their peak responses and bandwidths. (b) Which will sound more even across the bass register? (c) Which will sustain longer?

Solution to Exercise 10.13

(a) Peak response is proportional to Q, so the second is $9/4 = 2.25$ times higher at the peak. Bandwidth is f_0/Q, so the second's is $4/9 = 0.44$ times as wide.

(b) The first, with $Q = 4$. Its broad, low peak supports a wider range of notes roughly equally; the second boosts a narrow band strongly and leaves the notes either side comparatively weak.

(c) The second. A higher Q means less energy leaving per cycle, which is exactly Chapter 4's trade-off, and it means a quieter instrument.

Therefore the choice is between a loud, even guitar and a quieter, more colored one with longer sustain. Neither is correct; makers and players genuinely disagree, and this is what they are disagreeing about.

Exercise 10.14

(Challenging) A piano tuner stretches each octave by matching the lower note's 2nd partial. Take $B = 2.0 \times 10^{-4}$, constant. (a) How many cents is one octave stretched? (b) If four octaves are tuned upward in succession from the middle, what is the accumulated deviation at the top? (c) Compare with the ± 30 cents of a real Railsback curve, and say why the real answer is larger.

Solution to Exercise 10.14

(a) The stretch per octave is the amount the 2nd partial is sharp:

$$1200 \log_2 \sqrt{\frac{1+4B}{1+B}} = 600 \log_2 \left(\frac{1.00080}{1.00020} \right) = 600(0.000865) = 0.52 \text{ cents.} \tag{10.17}$$

(b) Four octaves: $4 \times 0.52 = 2.1$ cents.

(c) Two reasons the real figure is an order of magnitude larger.

First, B is **not** constant. It rises steeply toward both ends of the keyboard, as Figure 10.12 shows, so the top octaves are stretched far more than 0.5 cents each.

Second, tuners do not only match 2nd partials. Fifths, fourths, and double octaves are checked too, each involving higher partials which are sharper still, and the stretch a tuner settles on reflects all of them.

Therefore the single-partial calculation gives the right mechanism and the wrong magnitude, so the figure plots a measured curve rather than a computed one.

Chapter 11

Wind Instruments and Air-Column Resonance

Learning Objectives

By the end of this chapter, you should be able to:

- Determine the resonant frequencies of open and stopped cylindrical air columns, and explain the octave difference between them.
- Explain why a conical bore produces a complete harmonic series while a cylindrical bore closed at one end produces only odd harmonics.
- Apply the end correction to an open pipe, and explain why the effective length exceeds the physical length.
- Explain how an air jet at an edge produces and sustains oscillation in a flute or a flue organ pipe.
- Describe the operation of a reed as a pressure-controlled valve, and distinguish inward- and outward-striking reeds.
- Explain why the clarinet overblows a twelfth while the saxophone and oboe overblow an octave, in terms of bore shape.
- Explain how the lips of a brass player act as a reed, and how the player selects among the instrument's resonances.
- Explain why the resonances of a brass instrument are nearly, but not exactly, a harmonic series, and describe the roles of the mouthpiece, the bore flare, and the bell.
- Describe the four ways of changing the sounding length of a wind instrument – holes, keys, valves, and slides – and state the intonation compromise each involves.

Introduction

Wind instruments are organized differently from string instruments, and the difference comes first, before any details.

In a string instrument, the string decides the pitch. The body responds to whatever the string offers, and could be removed, leaving a quiet instrument, but one playing the same notes.

In a wind instrument, the **resonator decides the pitch**. The reed, the lips, or the air jet is a valve that will oscillate at almost any frequency if left to itself; what makes it oscillate at a *musical* frequency is feedback from the air column. Remove the bore from a clarinet and the reed produces a thin squawk at whatever frequency the player's lip happens to allow. Attach the bore and the reed plays what the bore tells it to.



Figure 11.1: The bore, keys, mouthpiece, and bell are a coupled acoustic system. Gisbert K, CC0.

This inverts the usual relationship. The air column is not an amplifier for the reed; the reed is a power supply for the air column.

Three questions organize the chapter. **What set of resonances does a given bore have?**, which turns out to depend on bore *shape* far more than on anything else. **How does the excitation lock onto them?**: the mechanisms differ interestingly between flutes, reeds, and lips. **How is the sounding length changed?**, by holes, keys, valves, or a slide, each with its own compromise.

11.1 Resonances of an Air Column

Pressure and Displacement Nodes

Chapter 3 established the boundary conditions; they are restated carefully here because they are the commonest source of confusion in the subject.

At a **closed** end the air cannot move. That is a **displacement node**. But it is exactly where the pressure swings hardest, since the air is being alternately compressed against the wall and pulled away from it; so it is a **pressure antinode**.

At an **open** end the air is free to move, and the pressure is held close to atmospheric by the enormous reservoir outside. That is a **displacement antinode** and a **pressure node**.

Any diagram of an air column must therefore say which quantity it is drawing, because the two pictures are each other's inverses. This book draws displacement unless it says otherwise.

The Open Pipe

A pipe open at both ends needs displacement antinodes at both, which requires a whole number of half-wavelengths:

$$f_n = n\,\frac{v}{2L}, \qquad n = 1, 2, 3, \ldots \tag{11.1}$$

A complete harmonic series. The flute, the recorder, and an open organ flue pipe all behave this way.

The Stopped Pipe and Its Odd Harmonics

A pipe closed at one end needs a node at that end and an antinode at the other, which are a quarter wavelength apart:

$$f_n = n\,\frac{v}{4L}, \qquad n = 1, 3, 5, \ldots \tag{11.2}$$

Two consequences, both audible. The fundamental is an **octave lower** for the same length. And **every even harmonic is missing**, which makes the tone hollow.

The clarinet is effectively a stopped pipe: the reed end is closed, the reed is a stiff barrier for most of the cycle, and the bell end is open.

11.2 Bore Shape

Cylinders and Cones

Now a result that surprises everyone the first time.

A **cone**, closed at the apex, has a *complete* harmonic series:

$$f_n = n\,\frac{v}{2L}, \qquad n = 1, 2, 3, \ldots \tag{11.3}$$

the same as an open cylinder, despite being closed at one end.

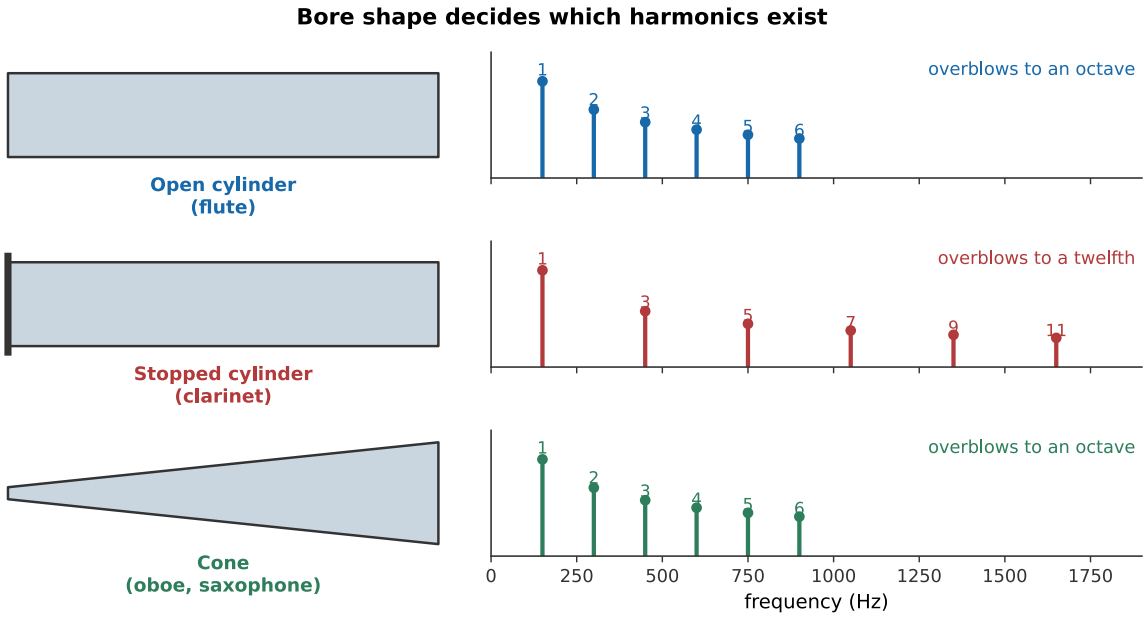


Figure 11.2: Three bores and the resonances each supports. The stopped cylinder is the odd one out: only odd harmonics, and therefore a hollow tone and an overblow to a twelfth rather than an octave. The cone, despite being closed at its apex, behaves like the open cylinder.

Why the Cone Restores the Even Harmonics

The physical reason is worth having, because it is not obvious.

In a cylinder the wave is **plane**: the same pressure across the whole cross-section, traveling straight down the tube. In a cone the wave is **spherical**, spreading from the apex, and a spherical wave's amplitude falls as $1/r$ as it expands.

Solving the wave equation in a cone gives pressure solutions of the form $\sin(kr)/r$. That function is zero at the apex regardless of k; so the apex's boundary condition costs nothing, and the resonances are set entirely by the open end. The result is the full series.

Put crudely: **a cone's closed end is not really a constraint**, because the spherical spreading already forces the pressure to behave correctly there.

The musical consequences are large. A saxophone and a clarinet both have single reeds and are both closed at the reed end. The saxophone is conical and the clarinet cylindrical, and they differ accordingly:

	Clarinet	Saxophone
Bore	Cylindrical	Conical
Harmonics	Odd only	All
Overblows to	A twelfth	An octave
Tone	Hollow, woody	Full, reedy
Length for a given pitch	Half	Full

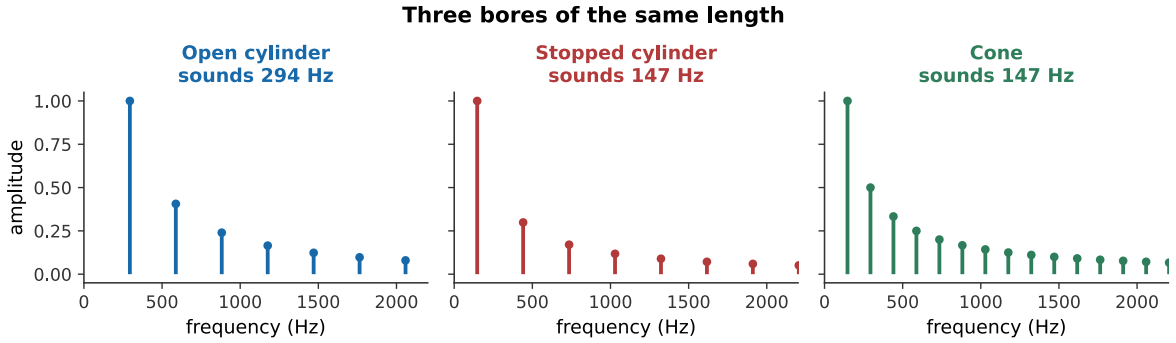


Figure 11.3: Three bores of the same physical length. Compare the second and third carefully: they sound the *same pitch*, both being closed at one end, and differ entirely in whether the even harmonics are there. *What you hear:*Three sustained tones. The first is bright and flute-like. The second is an octave lower and distinctly hollow. The third is the same pitch as the second but full and reedy rather than hollow. Listen: Open cylinder, Stopped cylinder, Cone

The End Correction

A pipe sounds as though it were slightly longer than it is, because the pressure node does not fall exactly at the opening: the air just outside is still being driven.

$$L_{\text{eff}} = L + 0.6\,r \quad \text{per open end.} \tag{11.4}$$

This is not a refinement to be mentioned and forgotten. An organ builder cutting pipes to length must include it, and because it depends on radius rather than on length, it matters more for short wide pipes than for long narrow ones: that is, more in the treble than in the bass, where a naive calculation would put the whole top of the instrument sharp.

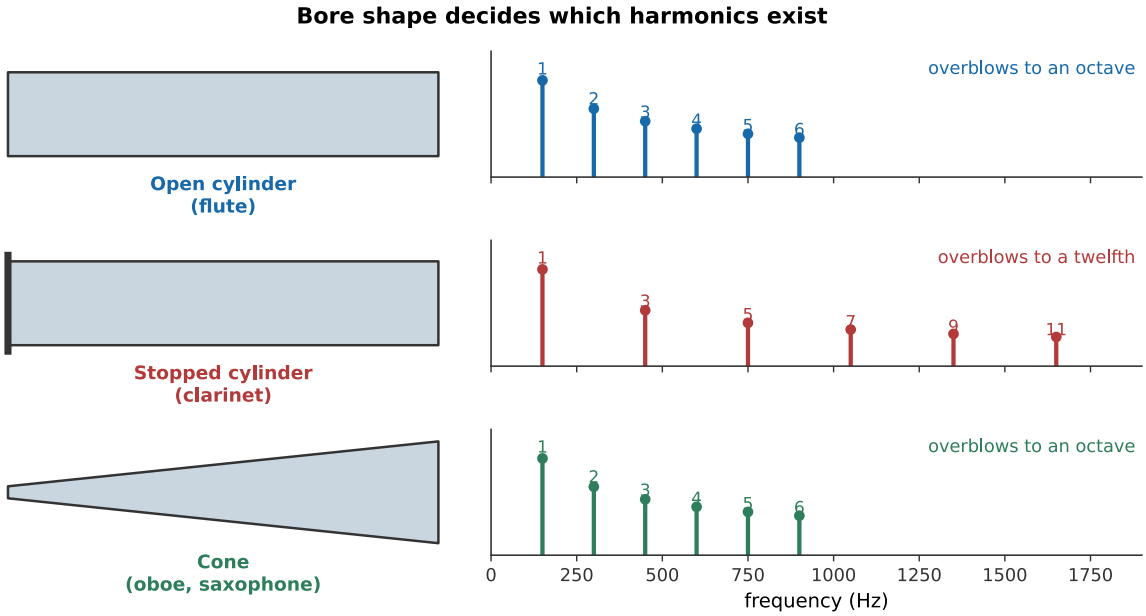


Figure 11.4: Overblowing. A cylindrical stopped pipe has no second harmonic to jump to, so it jumps to the third, a twelfth. A clarinet's register key therefore produces a twelfth, and a clarinetist must learn different fingerings for the two registers, while a saxophonist's fingerings simply repeat at the octave. *What you hear:*Two pairs of notes. In the first, the second note is a twelfth above the first, an octave and a fifth. In the second, the second note is exactly an octave above. Listen: Cylinder overblowing, Cone overblowing

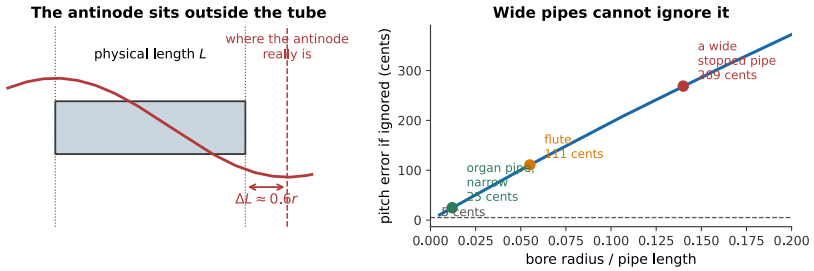


Figure 11.5: The end correction. **Left**: the antinode falls outside the tube. **Right**: what ignoring it costs. For a narrow organ pipe it is a few cents; for a wide, short pipe it can exceed a semitone.



Figure 11.6: Different lengths and constructions expose the geometric choices behind the simple air-column model. Saskii, English Wikipedia, public domain.

11.3 The Air Jet

Edge Tones

The flute family has no reed. Instead, a thin **jet** of air is directed across a sharp edge.

A jet crossing an edge is unstable. Small disturbances on it grow as it travels, so the jet flaps from one side of the edge to the other, alternately blowing into the pipe and outside it. On its own, this produces an **edge tone** whose frequency depends on the jet speed and the distance to the edge, and which glides continuously as the player blows harder.

Flutes and Recorders

Couple that jet to a pipe and the pipe takes control. A standing wave in the bore produces an oscillating flow at the mouth hole, which pushes the jet from side to side at the pipe's resonant frequency, which reinforces the standing wave. The jet locks to the pipe.

Two things follow that a player will recognize.

Blowing harder raises the pitch a little. The jet has a natural preference of its own, and forcing it away from that preference pulls the locked frequency slightly. Flutists correct constantly by rolling the instrument in or out, changing the effective mouth-hole geometry.

The edge tone is the noise you hear. A flute's characteristic breathiness is the jet's unlocked, turbulent component, and it is present in every note. Synthesizing a flute without it produces something that sounds like an organ.

Flue Organ Pipes

An organ flue pipe is the same mechanism made permanent. The jet speed is set by the wind pressure of the whole organ rather than by a player, and everything a flutist does with embouchure is instead built into the pipe's geometry at the time of voicing: the height of the mouth, the sharpness of the lip, the width of the windway.

An organ pipe therefore has exactly one tone color and one dynamic. An organ needs so many pipes for that reason: every note of every color needs its own, and a large instrument has five thousand of them.

11.4 Reed Instruments

The Reed as a Valve

A reed does not vibrate like a string, and treating it as one gets the physics wrong.

A reed is a **pressure-controlled valve**. The player supplies steady pressure; the reed opens and closes a gap; the rate of opening and closing determines the frequency of the puffs of air entering the bore.

The evidence that the bore is in charge is immediate: a clarinet reed's own natural frequency is a couple of kilohertz, and the instrument plays notes from 147 Hz upward. The reed is being driven far below its own resonance, and it goes where it is told.

The same reed therefore plays every note of the instrument, and changing the fingering changes the pitch without the player touching the reed.

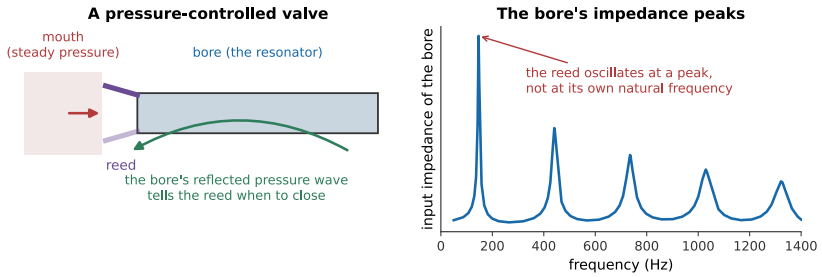


Figure 11.7: The feedback loop. The reed lets a puff of air into the bore; the bore reflects a pressure wave back; that returning pressure helps close the reed at the right moment. The reed oscillates at a frequency the **bore** chooses, not at its own natural frequency.

> **Note**
> A reed's behavior depends on whether increasing pressure tends to *close* it or *open* it. A clarinet's reed is **inward-striking**: mouth pressure pushes it toward closed, so it must be driven below its own resonance. A harmonium's or accordion's free reed is **outward-striking** and behaves quite differently, playing at a frequency close to its own. It is the inward-striking geometry that lets one reed serve a whole instrument.

The Clarinet and the Twelfth

The clarinet is a cylindrical stopped pipe, so §11.2 applies in full: odd harmonics, a tone that is hollow in the low register, and an overblow to a twelfth.

That twelfth is the clarinet's central practical problem. A twelfth spans nineteen semitones, so nineteen different fingerings are needed before the register key can be used at all, against eleven for an octave-overblowing instrument. This is the reason the clarinet has so many keys, why its fingering is so much harder to learn than the saxophone's, and why the awkward "throat" notes around the register break are the weakest on the instrument.

The compensation is that the clarinet gets a low register out of a short tube, and gets it with a distinctive hollow color that no other woodwind provides.

The Conical Reeds: Oboe, Bassoon, Saxophone

The oboe and bassoon use double reeds and the saxophone a single one, but acoustically the important thing they share is a **conical** bore. All three therefore produce complete harmonic series and overblow at the octave.

The conical bore also explains a structural point. A cone must come to a point to be acoustically complete, and a real instrument's reed sits where the point would be. The reed and its staple are designed so that their internal volume is approximately equal to that of the missing conical tip: an equivalent-volume substitution which, if it is wrong, throws the instrument's octaves out. Oboists trimming a reed are adjusting this, whether or not they describe it that way.
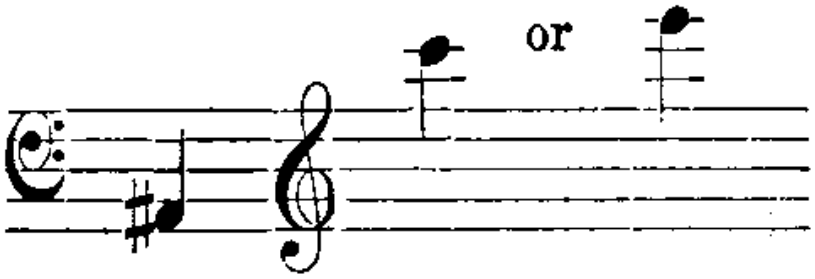


Figure 11.8: Changing the effective tube length changes the available resonance series. Victor Charles Mahillon and Kathleen Schlesinger, public domain.

11.5 Brass Instruments

The Lip Reed

A brass player's lips are a reed made of flesh, and the feedback mechanism is the same as §11.4's. The lips buzz; the bore reflects; the reflection controls the buzz.

The difference is that lips are **adjustable**. A player changes their tension and aperture continuously, and by doing so chooses which of the bore's resonances to lock onto. A trumpeter plays a scale on a fixed tube by changing nothing but their lips.

Mouthpiece, Bore, and Bell

Now a problem. A cylindrical tube closed at one end gives odd harmonics only, $1, 3, 5, \ldots$, which would make a brass instrument's playable notes an unmusical set. Yet a bugle plays a recognizable harmonic series.

The resolution is that a brass instrument is not a cylinder. It is a carefully shaped combination of three elements that between them force the resonances into line.

The **mouthpiece** is a small cavity with a constriction: a Helmholtz resonator (Chapter 4) with a resonance around 800 Hz. Its effect is to pull the upper resonances down into alignment, and to boost the output in the region where the ear is most sensitive.

The **bore** is mostly cylindrical in a trumpet, mostly conical in a horn, and the mixture determines how the resonances line up as well as the instrument's character.

The **bell** does two jobs, and they are opposite.

Below a **cut-off frequency** set by the flare, the bell reflects sound back into the bore, which is essential, because without reflection there is no standing wave and no resonance to lock onto. Above the cut-off, it radiates efficiently into the room, so brass instruments are loud and their high harmonics beam forward.

This split is audible. It is why a trumpet's low notes fill a hall while its high notes go where the bell points (Chapter 2), and why a player turning aside changes the sound so completely.

Making the Resonances Harmonic

The first resonance stubbornly refuses to join the series. A brass instrument's lowest resonance is typically well below where a true fundamental would sit, and it is essentially unusable.

Players nonetheless produce a note there, the **pedal tone**, and the mechanism is Chapter 8's missing fundamental: the lips buzz at the fundamental frequency, the higher resonances (2,



Figure 11.9: A physicist explains the lips as an adjustable oscillating valve and the air column as the resonator that selects its pitch. The demonstration shows the feedback loop behind the compact description above. Watch on YouTube: The Physics of Brass Instruments, Part 1: The Lips

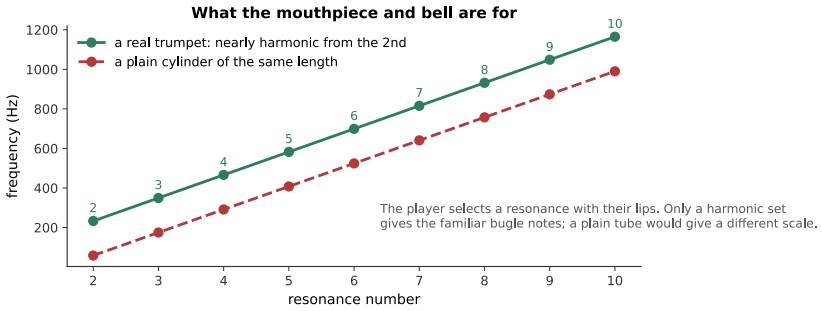


Figure 11.10: What the mouthpiece and bell are for. A plain cylindrical tube's resonances are an unusable set. A trumpet's are pulled into a harmonic series from the second resonance upward, and the player's lips then select among them.

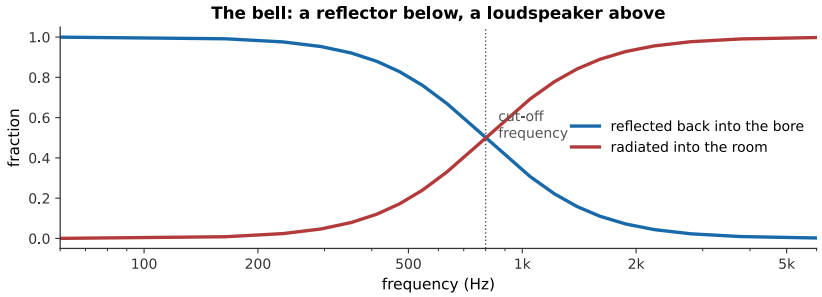


Figure 11.11: The bell as a frequency-dependent mirror. Below the cut-off it reflects almost everything, which is what keeps the standing wave alive. Above it, it radiates almost everything, which is what makes the instrument loud and directional.

3, 4, …) all support harmonics of it, and the ear supplies the fundamental that the instrument is barely radiating.

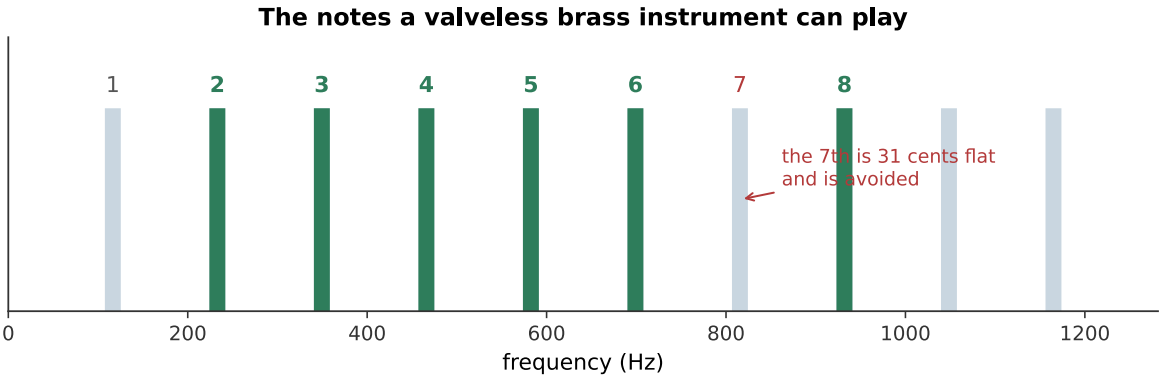


Figure 11.12: Resonances 2 through 8 of a fixed tube, the complete vocabulary of a valveless brass instrument. Every bugle call in every army is made from these notes, so they all sound alike. The 7th resonance is skipped: Chapter 9 showed it is 31 cents flat of any key. *What you hear:*A rising and falling sequence of eleven notes on a single brass tone, unmistakably a bugle call, using only the notes available without valves. Listen: Bugle

11.6 Changing the Length

Tone Holes and the Register Hole

A woodwind changes its sounding length by opening holes. Opening a hole creates a pressure node near it, effectively ending the tube there.

The approximation is imperfect. A hole is not a complete opening: some of the standing wave leaks past it and continues down the remaining tube, so the effective end is a little beyond the hole and depends on the hole's size. Large holes approximate a true open end better, but large holes are hard to cover with fingers: a tension that runs through the whole history of woodwind design and that keywork exists to resolve.



Figure 11.13: An air column is shaped by its physical length and by the openings that interrupt it. The Metropolitan Museum of Art, CC0.

Worked example: what a tone hole does to the pitch

The flute of Exercise 11.1 has a sounding length of 65.5 cm and a fundamental of 262 Hz. A tone hole is opened 49.1 cm from the embouchure hole, three-quarters of the way down the tube. Treating the open tone hole as a new open end, what note does the flute now sound, as an interval above the original?

The tube behaves as an open pipe of the new, shorter length:

$$f_1' = \frac{v}{2L'} = \frac{343 \text{ m/s}}{2(0.491 \text{ m})} = 349 \text{ Hz}. \tag{11.5}$$

The interval above the original 262 Hz is

$$1200 \log_2\left(\frac{349}{262}\right) = 1200 \log_2\left(\frac{4}{3}\right) = 498 \text{ cents}, \tag{11.6}$$

a perfect fourth, because 65.5/49.1 is very nearly 4/3.

Therefore a single tone hole, placed at three-quarters of the tube's length, raises the pitch a fourth: exactly the kind of arithmetic a woodwind maker uses when laying out a scale of holes along a bore. Real instruments depart from this idealization in the direction already noted above, some of the standing wave leaks past even an open hole, so the effective end sits a little beyond it, and the true pitch comes out slightly lower than this calculation predicts, more so for a small hole than a large one.

A **register hole** is a small hole placed near a pressure node of the desired higher mode. It barely disturbs that mode while spoiling the fundamental, so the instrument jumps up. Its placement is a compromise, because one hole must serve many notes, and this is the origin of the weak, stuffy notes near a clarinet's register break.

Keys and the Woodwind Compromise

Keywork exists to put holes where acoustics wants them rather than where fingers reach. The Boehm system, developed for the flute in the 1830s and adapted to the clarinet and saxophone, uses rods and levers so that one finger can close a distant or oversized hole.

The gain is intonation and evenness; the cost is mechanical complexity, and a modern clarinet has around twenty keys.

Valves, Slides, and Their Intonation Problems

Brass instruments change length by adding tube.

A **slide**, as on a trombone, adds a continuously variable length. It is acoustically perfect, any length is available, so any pitch is, and it is why a trombonist can play in just intonation with a string section.

Valves divert the air through fixed extra loops: conventionally the second lowers by a semitone, the first by a tone, and the third by a minor third. Combinations give the rest.

The combinations are systematically sharp, and the reason is arithmetic. Each valve is cut to lower the pitch of the *open* instrument by its interval. But using two valves together means the second is lengthening an already-lengthened tube, and lowering a longer tube by a fixed interval requires a *longer* addition than lowering the original. The fixed loop is now too short, and the note comes out sharp, by around 20 cents for two valves and 35 or more for three.

Players and makers address this in three ways: alternate fingerings, a movable slide on the first or third valve operated while playing, and compensating systems that route the air through extra tubing when certain combinations are used. None is a complete solution, so brass intonation is a skill rather than a property of the instrument.

11.7 Summary

- **In a wind instrument the resonator sets the pitch**, and the reed, lips, or air jet is a valve that locks to it. This inverts the string instrument's arrangement.
- A **closed end** is a displacement node and a pressure antinode; an **open end** is the reverse. Any diagram must say which it shows.
- **Open cylinder**: $f_n = nv/2L$, all harmonics. **Stopped cylinder**: $f_n = nv/4L$, odd harmonics only: an octave lower for the same length, hollow, and overblowing to a twelfth.
- **A cone closed at the apex has a complete harmonic series**, because the spherical solution $\sin(kr)/r$ vanishes at the apex for every k, so the closed end imposes no constraint. This is the whole difference between the clarinet and the saxophone.
- **The end correction** adds about $0.6r$ per open end, and matters most for short wide pipes.
- **An air jet at an edge** is unstable and locks to the pipe's resonance; the unlocked component is the breathiness, which is part of the sound rather than a defect.
- **A reed is a pressure-controlled valve** driven far below its own resonance. The bore's impedance peaks choose the frequency, so one reed plays every note.
- **A brass instrument's mouthpiece and bell force its resonances into a harmonic series** from the second upward. The bell reflects below its cut-off, essential for the standing wave, and radiates above it, which makes the instrument loud and directional.
- **The pedal tone** is a missing fundamental: the lips buzz at a frequency the instrument barely radiates, and the ear supplies it from the harmonics.
- **Valve combinations are systematically sharp**, because a fixed length added to an already-lengthened tube lowers the pitch by less than intended. Slides have no such problem.

11.8 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple choice.** Compared with an open cylindrical pipe of the same length, what are the fundamental and harmonic series of a stopped cylindrical pipe?
2. **True or false.** A cone closed at its apex has only odd harmonics, just like a stopped cylinder.
3. **Drag the words.** Complete the pressure and displacement node/antinode conditions at open and closed pipe ends.
4. **Fill in the blanks.** A clarinet's stopped-cylinder bore overblows to the _____, while a conical saxophone overblows to the _____.
5. **Mark the words.** Identify the two brass-instrument features that pull resonances toward a useful harmonic series: "The mouthpiece lowers upper resonances and the bell raises lower resonances."

11.9 Conceptual Questions

1. Explain why removing the bore from a clarinet stops it playing a musical note, and why removing the body from a violin does not.
2. State the boundary conditions at a closed and an open end for both displacement and pressure, and explain why a figure must say which it shows.
3. Explain why a clarinet overblows to a twelfth while a saxophone overblows to an octave, given that both are closed at the reed end.
4. Explain, without algebra, why the closed apex of a cone imposes no constraint on the resonant frequencies.
5. A flutist blows harder and the pitch rises slightly. Explain the mechanism, and say what the player does about it.
6. Explain why a brass instrument's bell must reflect some sound rather than radiating all of it.
7. A brass player produces a pedal tone at a frequency where the instrument has no usable resonance. Explain how.
8. Explain why pressing two valves together produces a note that is sharp, and name one remedy.

11.10 Problems

Exercise 11.1
(Straightforward) A flute behaves as an open cylinder. Its lowest note is 262 Hz at 20 °C. (a) What is its sounding length? (b) What are its next two resonances?

Solution to Exercise 11.1
(a) For an open pipe, $f_1 = v/2L$:

$$L = \frac{v}{2f_1} = \frac{343\ \text{m/s}}{2(262\ \text{Hz})} = 0.655\ \text{m}. \tag{11.7}$$

(b) All harmonics: 524 Hz and 786 Hz.

Therefore about 65.5 cm of sounding length, with resonances at 262, 524, 786 Hz: a complete harmonic series, so a flute overblows to the octave.

Exercise 11.2
(Moderate) A clarinet behaves as a stopped cylinder and its lowest note sounds 147 Hz. (a) What is its sounding length? (b) What are its next two resonances? (c) What interval is the first overblow?

Solution to Exercise 11.2
(a) For a stopped pipe, $f_1 = v/4L$:

$$L = \frac{343\ \text{m/s}}{4(147\ \text{Hz})} = 0.583\ \text{m}. \tag{11.8}$$

(b) Odd harmonics only: $3 \times 147 = 441$ Hz and $5 \times 147 = 735$ Hz.

(c) From 147 to 441 Hz is a ratio of 3:

$$1200 \log_2 3 = 1902\ \text{cents} = 19\ \text{semitones}, \tag{11.9}$$

a twelfth.

Therefore a clarinet gets its low register from 58 cm of tube where a flute would need 117 cm, and pays for it with a nineteen-semitone overblow.

Exercise 11.3
(Moderate) A pipe is 42.0 cm long and 2.2 cm in radius, open at both ends. (a) Find its fundamental ignoring end corrections. (b) Find it including them. (c) Express the difference in cents.

Solution to Exercise 11.3
(a) $f_1 = (343\ \text{m/s})/(2 \times 0.420\ \text{m}) = 408$ Hz.

(b) Two open ends, each adding $0.6r = 0.0132$ m:

$$L_{\text{eff}} = 0.420\ \text{m} + 2(0.0132\ \text{m}) = 0.446\ \text{m}, \qquad f_1 = \frac{343\ \text{m/s}}{2(0.446\ \text{m})} = 384\ \text{Hz}. \tag{11.10}$$

(c)

$$1200 \log_2(384/408) = -105\ \text{cents}. \tag{11.11}$$

Therefore ignoring the correction would put the prediction more than a semitone sharp. For a pipe of this proportion the correction is not optional.

Exercise 11.4

(Moderate) A saxophone and a clarinet both sound 147 Hz. (a) Compare their sounding lengths. (b) Compare the frequencies of their first four resonances. (c) Which is heavier to carry, and why is it worth it?

Solution to Exercise 11.4

(a) The clarinet is a stopped cylinder: $L = v/4f_1 = 0.583$ m. The saxophone is a cone, so $L = v/2f_1 = 1.17$ m, twice as long.

(b) Clarinet: 147, 441, 735, 1029 Hz (odd harmonics). Saxophone: 147, 294, 441, 588 Hz (all harmonics).

(c) The saxophone, being twice the tube for the same pitch. What it buys is a complete harmonic series, a fuller tone, and an octave overblow, which halves the number of distinct fingerings a player must learn.

Therefore the clarinet trades tone and fingering simplicity for compactness, and the saxophone trades compactness for both.

Exercise 11.5

(Moderate) A trumpet's tube is 1.48 m long. (a) Treating it as an open pipe, find the first four resonances. (b) Which of these can a player sound with the lips? (c) What are the corresponding musical intervals?

Solution to Exercise 11.5

(a) $f_1 = (343\text{ m/s})/(2 \times 1.48\text{ m}) = 116$ Hz, so the series is 116, 232, 348, 464 Hz.

(b) Resonances 2, 3, 4 and upward. The first is not properly supported by a real trumpet, for the reasons of §11.5, it lies below where the bell and mouthpiece pull the series into line.

(c) From 232 to 348 Hz is a fifth (3:2); 348 to 464 is a fourth (4:3); 232 to 464 is an octave.

Therefore the playable notes are the harmonic series from the second upward, and the intervals between successive resonances narrow as the player goes higher, high brass playing therefore requires fine lip control.

Exercise 11.6

(Moderate) A trumpet's second valve lowers the pitch by a semitone. (a) By what factor must the tube lengthen? (b) If the open tube is 1.48 m, how much tube must the valve add? (c) Repeat for the first valve, which lowers by a whole tone.

Solution to Exercise 11.6

(a) Frequency goes as $1/L$, so lowering by a semitone requires

$$L' = L \times 2^{1/12} = 1.05946\,L. \tag{11.12}$$

(b) The addition is

$$\Delta L = 1.48(1.05946 - 1) = 0.088\text{ m} = 8.8\text{ cm}. \tag{11.13}$$

(c) A whole tone needs $L' = L \times 2^{2/12} = 1.12246\,L$:

$$\Delta L = 1.48(0.12246) = 0.181\text{ m} = 18.1\text{ cm}. \tag{11.14}$$

Therefore about 8.8 cm and 18.1 cm of extra tubing. Note that the second is more than twice the first, which is the seed of the next problem.

Exercise 11.7
(Challenging) Using the trumpet of Exercise 11.6, a player presses valves 1 and 2 together, intending a minor third. (a) What total length results? (b) What length would a true minor third need? (c) How sharp is the note, in cents?

Solution to Exercise 11.7
(a) The two additions simply sum:

$$L = 1.48 + 0.088 + 0.181 = 1.749 \text{ m}. \tag{11.15}$$

(b) A true minor third is three semitones:

$$L_{\text{true}} = 1.48 \times 2^{3/12} = 1.48 \times 1.18921 = 1.760 \text{ m}. \tag{11.16}$$

(c) The tube is short by 0.011 m, so the pitch is sharp by

$$1200 \log_2\!\left(\frac{1.760}{1.749}\right) = 10.9 \text{ cents}. \tag{11.17}$$

Therefore the combination is about 11 cents sharp. The cause is exactly as §11.6 describes: valve 1's loop was cut to lower the *open* tube by a tone, but here it must lower a tube already lengthened by valve 2, and a fixed length does less lowering on a longer tube.

Exercise 11.8
(Straightforward) A brass instrument's bell has a cut-off frequency of 1.5 kHz. (a) What happens to a 400 Hz component reaching the bell? (b) A 4 kHz component? (c) Explain why both behaviors are necessary.

Solution to Exercise 11.8
(a) 400 Hz is well below cut-off, so almost all of it is **reflected** back into the bore.

(b) 4 kHz is well above, so almost all of it is **radiated** into the room.

(c) Reflection is what sustains the standing wave: without a returning pressure wave there is no resonance for the lips to lock onto, and no note. Radiation is what the audience hears. An instrument that reflected everything would resonate beautifully and be inaudible; one that radiated everything would be loud for an instant and then have nothing left to oscillate with.

Therefore the bell must do both, and the cut-off is where it changes its mind, which is also why brass instruments sound so different on and off axis, since the radiated high components are the directional ones.

Exercise 11.9
(Moderate) An organ builder must make a stopped pipe sounding 65.4 Hz at 15 °C, with a bore radius of 5.0 cm. (a) Find the speed of sound. (b) Find the required effective length. (c) Find the physical length to cut.

Solution to Exercise 11.9

(a) $v = 331.3 + 0.606(15) = 340.4$ m/s.

(b) For a stopped pipe:

$$L_{\text{eff}} = \frac{v}{4f_1} = \frac{340.4}{4(65.4)} = 1.301 \text{ m}. \tag{11.18}$$

(c) One open end, so one correction of $0.6r = 0.030$ m:

$$L = 1.301 - 0.030 = 1.271 \text{ m}. \tag{11.19}$$

Therefore cut 127.1 cm. Cutting the uncorrected 130.1 cm would give an effective length of 133.1 cm and a pitch of 63.9 Hz, about 40 cents flat, which on an organ is unusable.

Exercise 11.10

(Challenging) A clarinet's register key must be placed near a pressure node of the third harmonic. The instrument's sounding length is 58.3 cm and it is stopped at the reed end. (a) Where along the tube is that node? (b) Explain why one hole cannot serve every note perfectly.

Solution to Exercise 11.10

(a) The third mode of a stopped pipe has $3\lambda/4 = L$, so $\lambda = 4L/3 = 0.777$ m. Pressure nodes occur at odd quarter-wavelengths from the closed end:

$$x = \frac{\lambda}{4} = 0.194 \text{ m}, \tag{11.20}$$

about 19.4 cm from the reed. (The next is at $3\lambda/4 = 58.3$ cm, the open end.)

(b) Because the sounding length changes with every fingering. As tone holes open, the effective tube shortens and the position of the third mode's pressure node moves with it, but the register hole is drilled once, in one place. It is therefore correctly placed for only one note and progressively wrong for the others.

Therefore the register break is inherently a compromise, and the notes furthest from the hole's ideal position are the stuffy, unreliable ones every clarinetist knows.

Exercise 11.11

(Challenging) Show that a cone closed at its apex has resonances at $f_n = nv/2L$, using the fact that pressure solutions in a cone have the form $p(r) \propto \sin(kr)/r$.

Solution to Exercise 11.11

Take the apex at $r = 0$ and the open end at $r = L$.

At the apex. The solution $p \propto \sin(kr)/r$ is finite there, since $\sin(kr)/r \to k$ as $r \to 0$. There is no condition to impose: the closed apex is automatically satisfied for *every* k, which is the whole point.

At the open end. The pressure must vanish:

$$\frac{\sin(kL)}{L} = 0 \;\Rightarrow\; \sin(kL) = 0 \;\Rightarrow\; kL = n\pi, \tag{11.21}$$

for $n = 1, 2, 3, \ldots$

Since $k = 2\pi f/v$:

$$\frac{2\pi f_n}{v} L = n\pi \qquad \Rightarrow \qquad f_n = \frac{nv}{2L}. \tag{11.22}$$

Therefore the cone gives every integer n, a complete harmonic series: the same as an open cylinder. Compare the stopped cylinder, where the closed end *does* impose a condition ($\mathrm{d}p/\mathrm{d}x = 0$) and halves the available set.

Exercise 11.12

(Moderate) A trumpeter plays a pedal tone. The instrument's resonances are at 232, 348, 464, 580 Hz. (a) What is the common fundamental of these? (b) Does the instrument have a strong resonance there? (c) Explain what the listener hears and why.

Solution to Exercise 11.12

(a) The spacing is 116 Hz and each listed frequency is a multiple of it, so the fundamental is 116 Hz.

(b) No. These are resonances 2, 3, 4, 5, and the first resonance of a real trumpet lies well away from 116 Hz, it is the one the bell and mouthpiece fail to pull into line.

(c) The player's lips buzz at 116 Hz. Each of its harmonics, 232, 348, 464, …, coincides with a strong resonance and is supported, so the instrument radiates a full harmonic series on 116 Hz while radiating almost nothing at 116 Hz itself. The ear then does what Chapter 8 describes and reports the fundamental.

Therefore the pedal tone is a missing fundamental, and a spectrum analyzer pointed at the bell would show very little energy at the pitch everyone in the room is hearing.

Exercise 11.13

(Moderate) A flute is tuned at 21 °C and played outdoors at 8 °C. (a) How far flat does it go? (b) The player pushes the head joint in to shorten the instrument. If the sounding length was 65.5 cm, by how much must they shorten it?

Solution to Exercise 11.13

(a) The speeds are $v(21) = 344.0$ m/s and $v(8) = 336.1$ m/s. Pitch follows the speed:

$$1200 \log_2\!\left(\frac{336.1}{344.0}\right) = -40 \text{ cents}. \tag{11.23}$$

(b) To recover 40 cents the length must fall by the same ratio:

$$L' = 0.655 \times \frac{336.1}{344.0} = 0.640 \text{ m}, \tag{11.24}$$

a shortening of 1.5 cm.

Therefore the flute goes 40 cents flat and needs about 15 mm of head-joint adjustment, which is a large movement, and more than the joint allows on some instruments. Outdoor performance in the cold is genuinely difficult for wind players for that reason.

Exercise 11.14

(Moderate) Compare a trombone and a valved trumpet as instruments for playing in just intonation with a string quartet. (a) Which is better suited, and why? (b) What must the trumpeter do instead? (c) Relate your answer to Chapter 9.

Solution to Exercise 11.14

(a) The trombone. Its slide provides a continuously variable tube length, so any pitch whatever is available: the player can place a third 14 cents flat of equal temperament to match a string section without effort.

(b) The trumpeter has only fixed valve combinations, each of which gives a pitch determined by the tubing. They must bend the note with their lips, "lipping" it up or down, which works over perhaps ± 30 cents but costs tone quality and stability, or use an alternate fingering, or operate a movable slide on the first or third valve while playing.

(c) Chapter 9 divided instruments into fixed-pitch and flexible-pitch, and noted that the compromise always falls on the flexible ones. A trombone is fully flexible and belongs with the strings and voices. A valved trumpet is *partly* fixed, which puts it in an awkward middle position: flexible enough to be expected to adjust, constrained enough that adjusting is work.

Chapter 12

Percussion: Membranes, Bars, and Plates

Learning Objectives

By the end of this chapter, you should be able to:

- Explain why the modes of a two-dimensional vibrator are generally inharmonic, and why this makes most percussion instruments indefinite in pitch.
- Describe the transverse modes of a free-free bar, and state the frequencies of the first few relative to the fundamental.
- Explain how undercutting a marimba bar brings its second mode into a harmonic relation with its first, and what that does to the perceived pitch.
- Describe the mode shapes of a circular membrane, identify nodal diameters and circles, and explain why an ideal drumhead has no definite pitch.
- Explain how the air loading of a kettle turns the timpani's inharmonic modes into a near-harmonic set, and identify which mode sounds the pitch.
- Describe the modes of a flat plate, interpret a Chladni pattern, and explain what the sand reveals.
- Explain why cymbals and gongs produce a dense, noise-like spectrum, and describe the nonlinear energy cascade responsible.
- Describe the modes of a church bell and the tuning practice that makes the hum, prime, tierce, quint, and nominal fall in fixed relations.
- Explain what it means to tune an inharmonic instrument, and identify the strategies used for bars, kettles, and bells.

Introduction

Every instrument so far has had a harmonic spectrum. Strings and air columns are one-dimensional, and their boundary conditions force their modes into exact whole-number ratios; so they have clear pitches, their tones blend, and Chapter 9's tuning systems are built out of small integers.

Percussion is different, and the difference is not a matter of degree.

A drumhead, a metal bar, a plate, a bell: these are two-dimensional or three-dimensional vibrators, and their modes fall in ratios like $1 : 1.59 : 2.14 : 2.30 : 2.65$. Nothing forces those into whole numbers, and they are not. The ear, presented with such a set, cannot find



Figure 12.1: Percussion instruments make mode shapes physical: a stretched membrane, a bowl, and an adjustable boundary all contribute to the sound. The Metropolitan Museum of Art, CC0.

a fundamental that explains it (Chapter 8), and reports a sound with a rough pitch region rather than a note.

Percussion therefore divides so sharply into instruments of **definite** and **indefinite** pitch, and the definite ones, timpani, marimba, tuned bells, are the result of considerable ingenuity. Each is an object that has been *made* to have a pitch, by moving its modes into place. This chapter is largely about how.

It is also where the book's model starts to strain. Chapter 5's spectrum assumed periodicity; a cymbal is not periodic. Chapter 8's consonance assumed harmonic partials; a gamelan's are not. The physics still works, it is just that the answers stop looking like music theory.

12.1 Why Percussion Is Different

One Dimension Versus Two

A string is a one-dimensional system, and its modes are indexed by a single number n. The boundary conditions, zero at both ends, force $\lambda_n = 2L/n$ and hence $f_n = nf_1$. The whole-number ratios come from counting half-wavelengths along one direction.

A membrane is two-dimensional, and a mode must satisfy a boundary condition around a whole *curve* rather than at two points. Its modes are indexed by two numbers, and the frequencies are set by the zeros of **Bessel functions** rather than by the zeros of a sine. Bessel zeros are irrational numbers with no arithmetic relationship to each other.

A bar is one-dimensional but obeys a different equation. A string's restoring force comes from **tension**; a bar's comes from **bending stiffness**, and the resulting equation has a fourth derivative rather than a second. Its modes come out in the ratios

$$1 : 2.756 : 5.404 : 8.933 : \ldots \tag{12.1}$$

which are the squares of the roots of $\cos x \cosh x = 1$, again, no whole numbers anywhere.

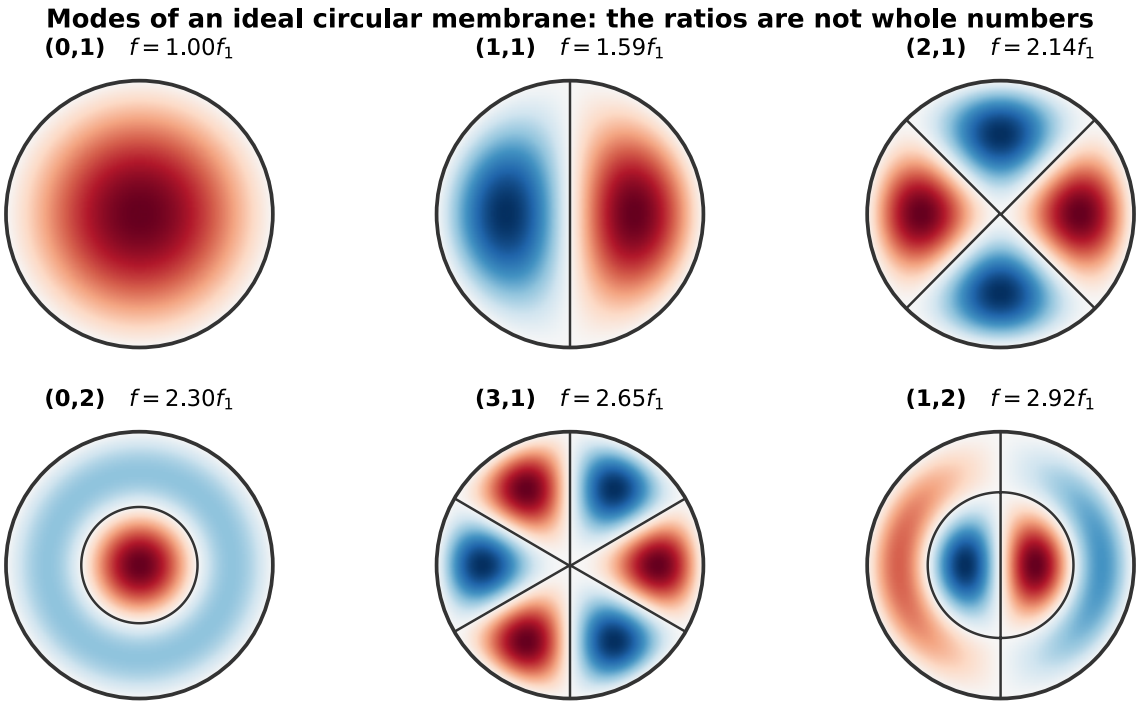


Figure 12.2: The first six modes of an ideal circular membrane. Each is labeled (m, n) with m nodal **diameters** and $n-1$ nodal **circles** inside the rim. The frequency ratios, 1.00, 1.59, 2.14, 2.30, 2.65, 2.92, are not whole numbers, are not close to whole numbers, and are not multiples of anything.

Inharmonic Modes and Indefinite Pitch

Chapter 8 explained what the ear does with a set of partials: it looks for a fundamental of which all are harmonics. Given $200, 300, 400, 500$ Hz it finds 100 Hz and reports it, even if nothing is there.

Given $200, 318, 428, 460, 530$ Hz it finds nothing. There is no fundamental that explains this set, and the pattern-matching mechanism has no answer to return. The listener hears a sound with a *center of gravity*, they can say it is high or low, and can tell one drum from another, but not a note. They cannot sing it back, and it does not participate in harmony.

Strike Tone and Decay

Two more features distinguish percussion.

The strike transient is a large part of the sound. Chapter 5 showed that attack carries an instrument's identity; for percussion it may carry most of it. The initial contact is broadband noise, a click, and removing it makes a struck bar sound like an organ.

Different partials decay at very different rates. High modes lose energy faster, so the spectrum thins as the sound dies. A struck bell begins as a clangorous mixture and ends as a nearly pure hum tone, and the change is one of the most recognizable things about it.

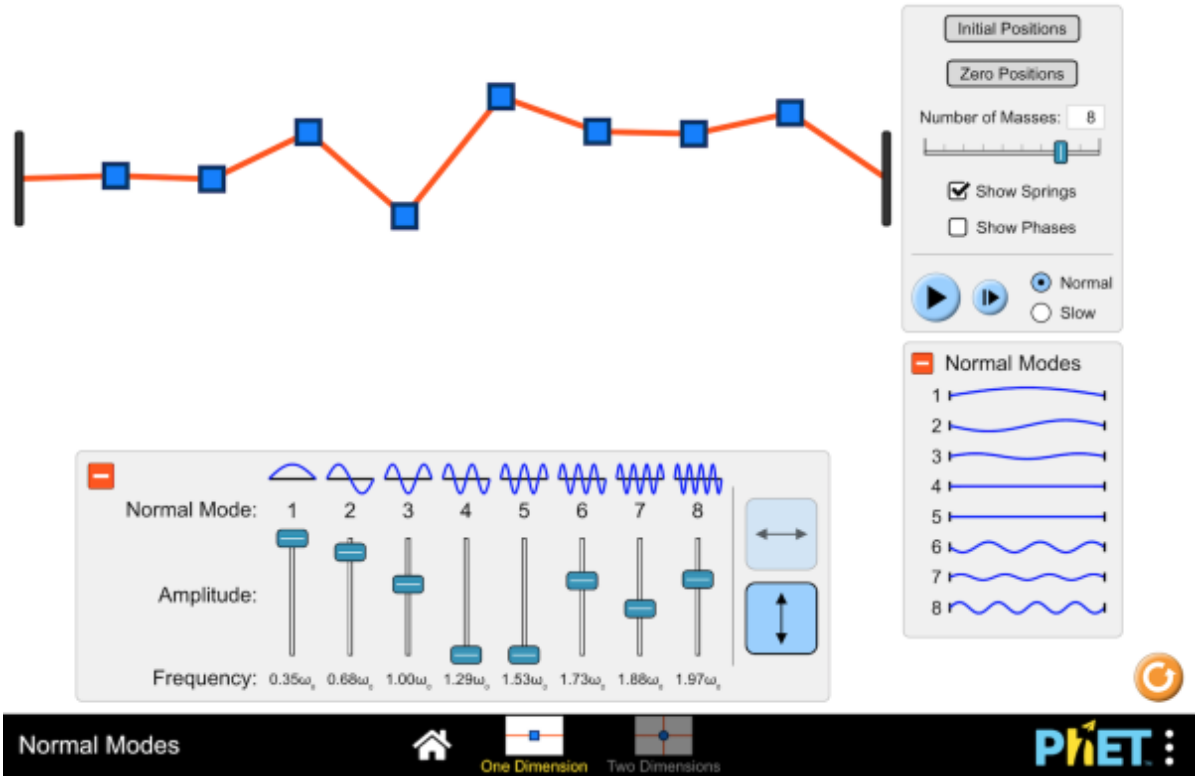


Figure 12.3: Switch to the two-dimensional screen. The modes of a rectangular array are indexed by two numbers, and their frequencies are emphatically not whole-number multiples of the lowest, which is this chapter's opening claim, in the simplest two-dimensional system there is. Interactive simulation: Normal Modes

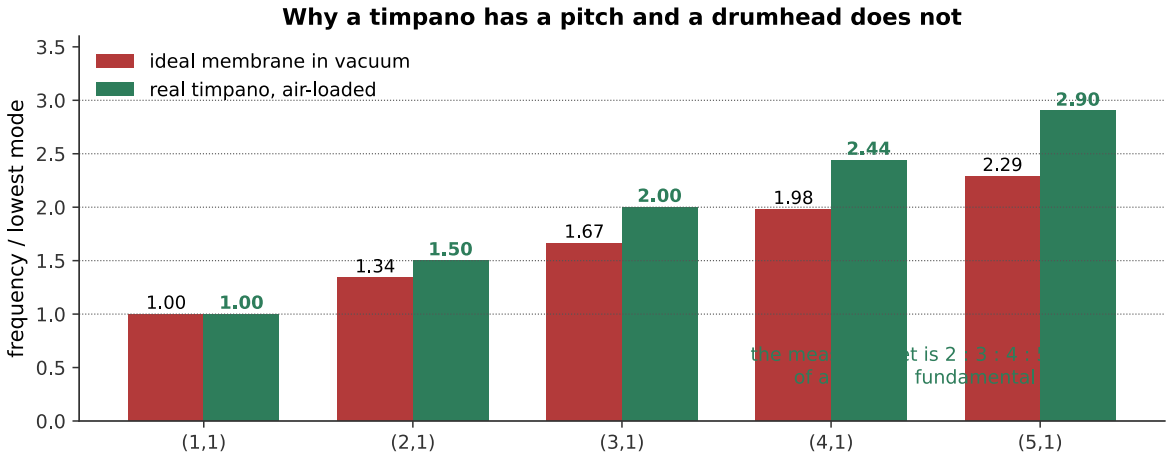


Figure 12.4: An ideal membrane and a modeled kettledrum, struck the same way. The difference is entirely in where the partials sit, and §12.3 explains how the kettle moves them. *What you hear:*Two struck drum sounds. The first is a dull thud with no identifiable note in it. The second is recognizably a timpano and has a clear pitch that could be sung back. Listen: Ideal membrane, Modeled timpano

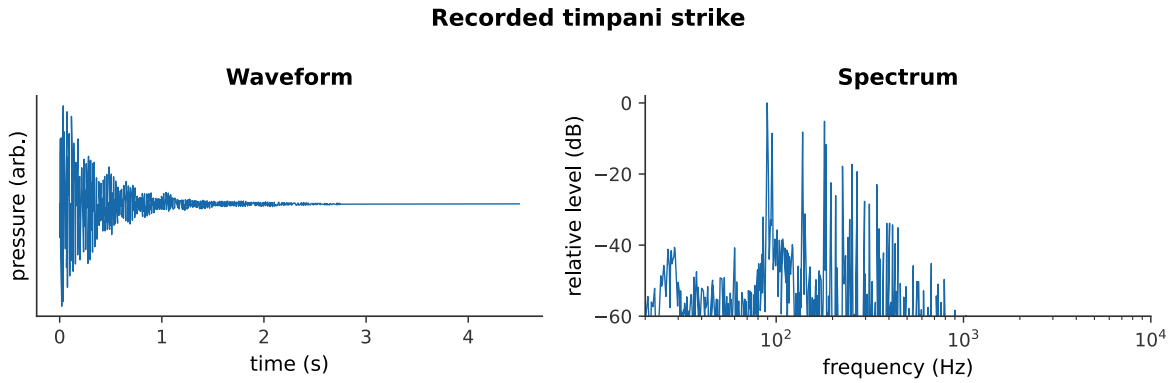


Figure 12.5: A recorded timpani. The synthesized comparison above isolates the shifted mode ratios; this CC0 recording restores everything the model leaves out: mallet noise, room sound, small mistunings, and interacting decays. *What you hear:*A real timpani strike begins with a short mallet impact, settles into a clearly pitched low ring, and decays for several seconds. Listen: Real Timpani

12.2 Bars

Modes of a Free-Free Bar

A xylophone or marimba bar is supported at two points and otherwise free. Its transverse modes are in the ratios $1 : 2.756 : 5.404 : \ldots$, and the second is the problem: 2.756 is close to neither 2 (an octave) nor 3 (a twelfth), and it clashes with the fundamental.

The supports are placed at the **nodes** of the fundamental mode, about 22% of the length from each end, so that holding the bar does not damp the note it is supposed to produce. This is the same principle as touching a string at a node (Chapter 3).

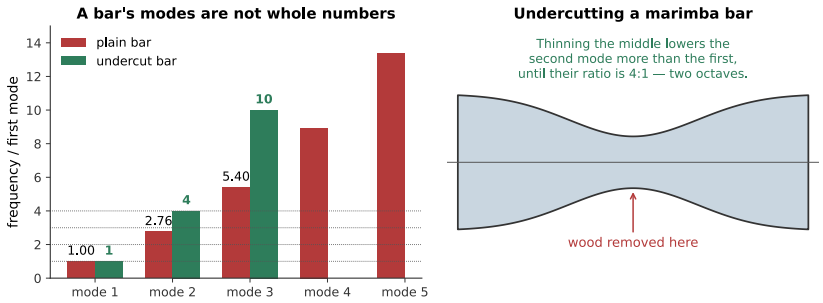


Figure 12.6: Left: a plain bar's modes against an undercut bar's. Right: where the wood comes off. Thinning the middle is not decoration, it is what makes a marimba a pitched instrument.

Undercutting: Xylophone, Marimba, Vibraphone

The fix is to remove material from the underside of the bar, and the reasoning is elegant.

The fundamental has its maximum displacement, its antinode, at the center of the bar. The second mode has a *node* there. Removing material from the center therefore weakens the bar where the fundamental is bending hardest and where the second mode is barely bending at all, so it lowers the fundamental far more than it lowers the second mode.

Remove enough and the ratio 2.756 can be driven up to 4.0, two octaves, an exact harmonic. Continue and the third mode can be brought to about 10.

The amount of undercutting distinguishes the instruments:

Instrument	Second mode tuned to	Character
Xylophone	$3f_1$ (a twelfth)	Bright, sharp, penetrating
Marimba	$4f_1$ (two octaves)	Warm, full, mellow
Vibraphone	$4f_1$	Mellow, sustained (aluminum, long decay)

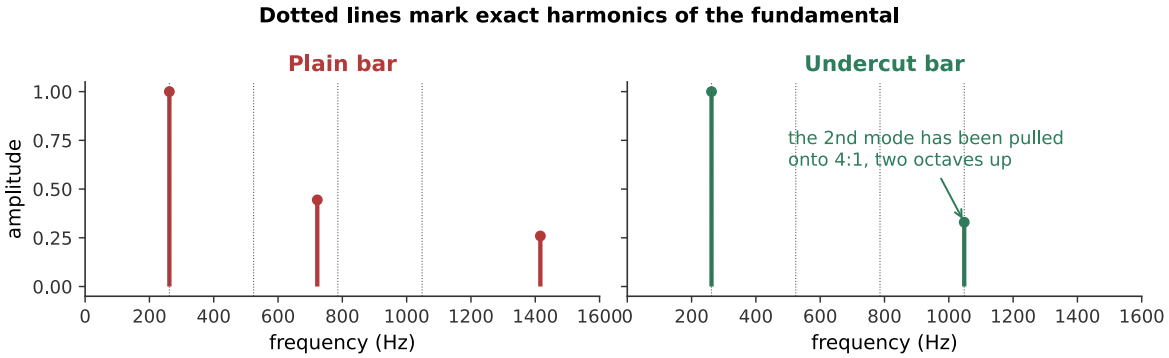


Figure 12.7: The same fundamental, with the second mode at 2.76 and at 4.0 times it. Moving one partial is enough to turn a clang into a note. *What you hear:*Two struck bars at the same fundamental. The first is clangorous and hard to assign a note to; the second has a clear, warm pitch. Listen: Plain bar, Undercut bar

Struck Bars That Are Not Undercut

A glockenspiel's bars are not undercut, and its steel bars keep the 2.756 ratio.

It works anyway, for a reason worth noticing: the glockenspiel plays **high**. Its fundamentals are above 2 kHz, so the inharmonic second mode lands above 5 kHz: a region where Chapter 8 showed pitch perception is poor and where the ear largely stops trying to match patterns. The partial is audible as brightness rather than as a clashing pitch.

Tubular bells exploit the same loophole from the other direction: their modes are inharmonic, but modes 4, 5, and 6 happen to fall close to the ratios $2:3:4$, and the ear constructs a **missing fundamental** from them. A tubular bell's perceived pitch is an octave below its lowest strong partial and corresponds to no mode of the tube at all.

12.3 Membranes

Modes of a Circular Membrane

The modes of an ideal circular membrane, fixed at the rim, are

$$f_{mn} = \frac{\alpha_{mn}}{2\pi a}\sqrt{\frac{T}{\sigma}}, \tag{12.2}$$

where a is the radius, T the tension per unit length, σ the mass per unit area, and α_{mn} the nth zero of the Bessel function J_m.

The ratios were listed in §12.1, and none of them is anything.

Worked example: the fundamental of a drumhead

A snare drum's batter head has radius $a = 0.178$ m (a 14-inch drum), areal mass density $\sigma = 0.20$ kg/m^2, and is tensioned to $T = 3200$ N/m. Using $\alpha_{01} = 2.405$, find its $(0,1)$ fundamental.

$$f_{01} = \frac{\alpha_{01}}{2\pi a}\sqrt{\frac{T}{\sigma}} = \frac{2.405}{2\pi(0.178\ \text{m})}\sqrt{\frac{3200\ \text{N/m}}{0.20\ \text{kg/m}^2}} = (2.151\ \text{m}^{-1})(126.5\ \text{m/s}) = 272\ \text{Hz}. \quad (12.3)$$

The next mode up, $(1,1)$, sits at 1.59 times this, about 432 Hz, an interval of

$$1200\log_2(1.59) = 804\ \text{cents}. \quad (12.4)$$

The nearest ratio of small whole numbers is 8:5 (813.7 cents), itself an unremarkable interval, not a number that organizes a scale, and even that near-miss does not extend to the modes above it. A snare drum, unlike a timpano, has no enclosed kettle of air to pull this family of ratios toward consonance, so its head produces this pattern undisguised: a perfectly well-defined f_{01}, and no definite pitch at all.

Nodal Diameters and Nodal Circles

The index m counts **nodal diameters**, straight lines across the head that do not move, and $n-1$ counts **nodal circles**.

This has a practical consequence percussionists use constantly. Striking at the center excites only the modes with no nodal diameter, $(0,1)$, $(0,2)$, and so on, because every mode with $m \geq 1$ has a node there. Striking near the rim excites the $(m,1)$ family strongly.

Timpanists strike about a quarter of the way in from the rim, and §12.3 explains why that is exactly right.

The Timpani: How the Kettle Fixes the Pitch

A timpano is a membrane with a pitch, and it gets one from three things at once.

The air in the kettle loads the head. The enclosed air resists being compressed, which adds stiffness, and it has mass, which the head must drag along. Crucially the loading is *different for different modes*: the $(0,1)$ mode moves the whole head in the same direction and compresses the kettle air strongly, so it is stiffened a great deal. The $(1,1)$ mode moves one half up while the other goes down, so the air merely sloshes sideways and is barely compressed at all.

The result is that the modes move by different amounts, and they move into a near-harmonic set.

The pitch heard is a missing fundamental. Multiply the measured set by two and it reads $2:3:4:5:6$, harmonics 2 through 6 of a note an octave below the $(1,1)$ mode. Exactly as with the brass pedal tone of Chapter 11, the ear supplies what the instrument does not radiate.

The $(0,1)$ mode, which moves the whole head as a piston, is the one mode that does *not* fit this series. It is also the loudest radiator, because a piston pushes far more air than a mode with nodal diameters. So a timpanist strikes about a quarter of the way in, near a node of

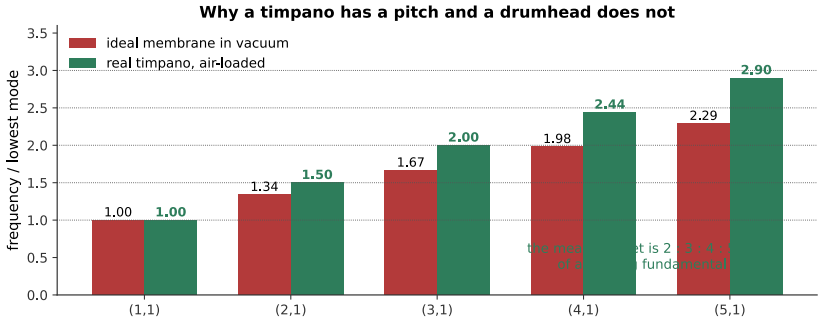


Figure 12.8: Ideal against real. The air-loaded timpani modes sit very close to the ratios $1 : 1.5 : 2 : 2.5 : 3$, which is $2 : 3 : 4 : 5 : 6$ of a fundamental one octave below the lowest mode. The ear finds that fundamental and reports it.

$(0, 1)$, suppressing the one mode that would spoil the pitch, and favoring the ones that make it.

12.4 Plates, Shells, and Bells

Flat Plates and Chladni Patterns

A flat plate has modes indexed by two numbers, like a membrane, but stiffness-controlled, like a bar. Their ratios are inharmonic and depend on the plate's shape and boundary conditions.

The classic way to see them is **Chladni's** method, published in 1787: scatter sand on the plate and excite a mode, and the sand bounces away from the moving regions and collects along the nodal lines.

Cymbals, Gongs, and the Nonlinear Cascade

A cymbal's modes are so numerous and so closely spaced that by a few hundred hertz the ear cannot resolve them. Its spectrum is effectively continuous, and it is heard as noise.

Something further happens when a cymbal is struck hard, and it is genuinely different physics. The plate's deflection becomes large enough that the restoring force is no longer proportional to displacement, the small-amplitude assumption of Chapter 1 fails, and the modes stop being independent. Energy fed into low modes **cascades** into higher ones, spreading across the spectrum over a few tenths of a second.

A crash cymbal therefore *blooms*: it does not simply start loud and decay, it starts with a limited spectrum and then fills out, brightening as it goes. It is one of the very few places in this book where superposition genuinely fails, and it is audible.

The Church Bell and Its Five Tuned Partials

A bell is the most remarkable object in this chapter: a three-dimensional shell whose modes have been forced, by centuries of trial and refinement, into deliberate musical relationships.

A tuned bell has five named partials:

Two things are worth drawing out.

The strike note is not a partial. The pitch a listener assigns to a bell corresponds to the prime, but it is largely constructed by the ear from the nominal, the quint, and higher



Figure 12.9: Physics Girl excites successive resonances of a Chladni plate. Sand leaves the moving regions and gathers on nodal lines, making each two-dimensional mode visible as a different geometric pattern. Watch on YouTube: Singing Plates: Standing Waves on Chladni Plates

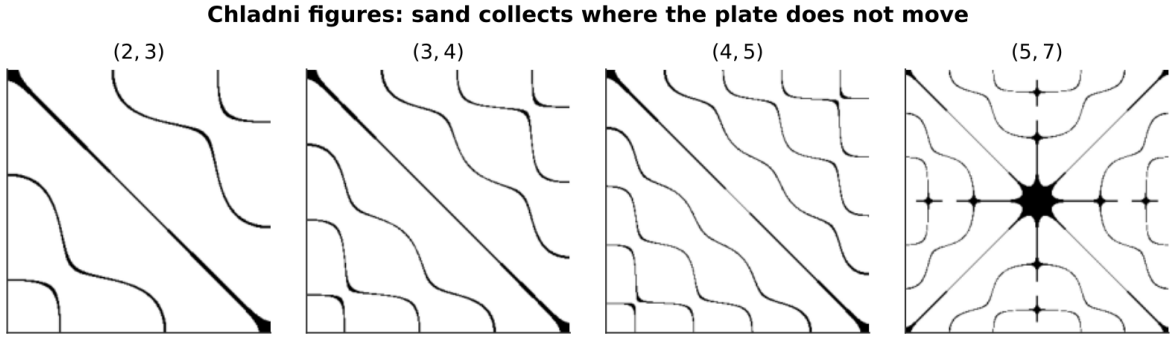


Figure 12.10: Select different resonant frequencies and watch the particles collect along the lines that do not move. Each pattern belongs to one mode, and the number of lines generally rises with the mode number. Violin makers still use the method to check the symmetry of a top or back plate before assembly; an asymmetric pattern means the plate is thicker on one side. Interactive simulation: Resonance — Chladni patterns

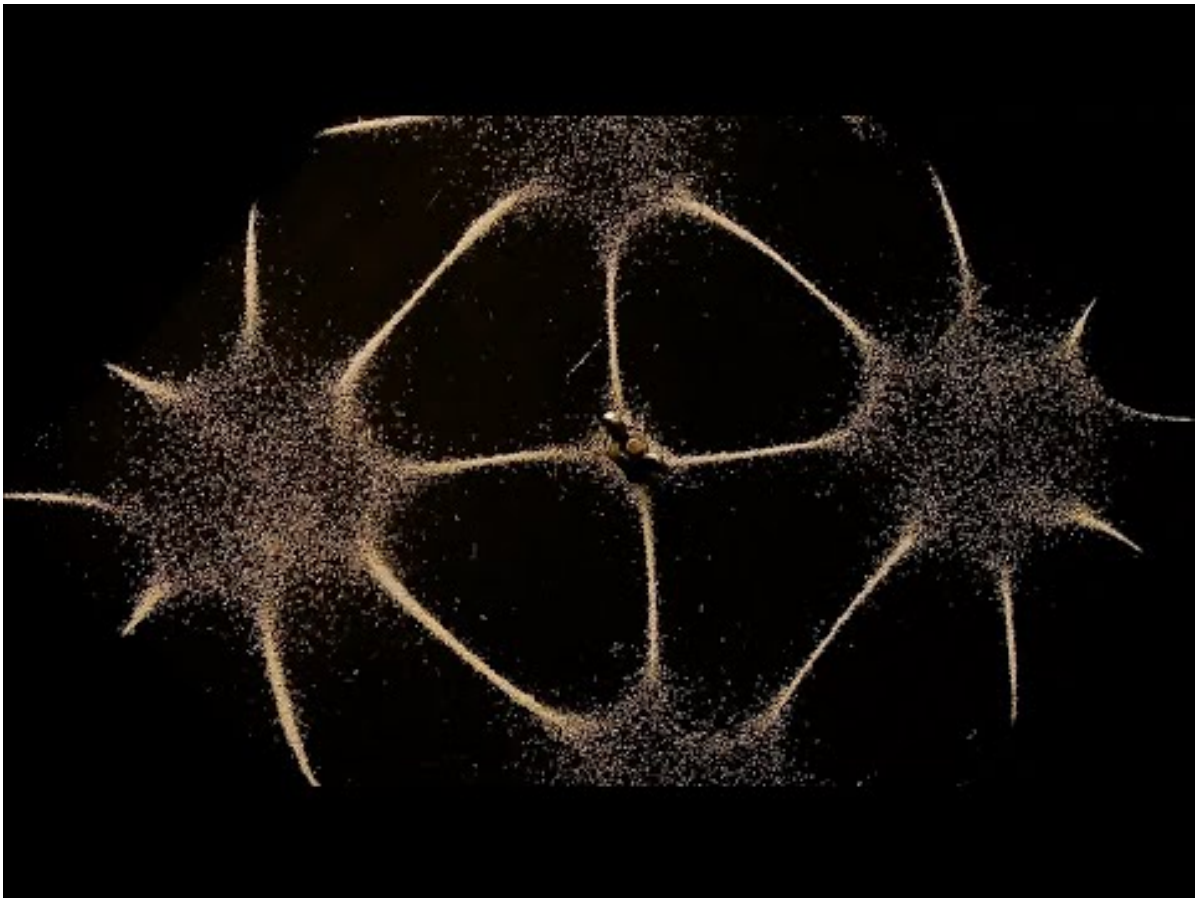


Figure 12.11: BBC Earth Science gives a slower explanatory view of the same method, showing why grains migrate away from vibrating regions and settle along the lines that do not move. Watch on YouTube: Seeing Sound With Sand

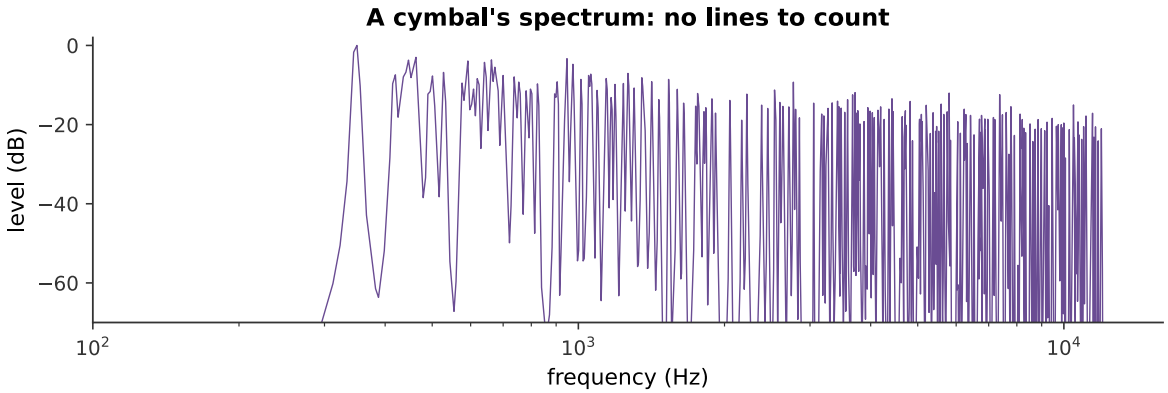


Figure 12.12: A cymbal-like sound built from hundreds of randomly spaced partials. There are no lines to count in the spectrum, and no fundamental for the ear to find, which is the definition of indefinite pitch. *What you hear:*A metallic crash that swells slightly after the initial strike and then decays over several seconds, with no pitch in it at all. Listen: Cymbal

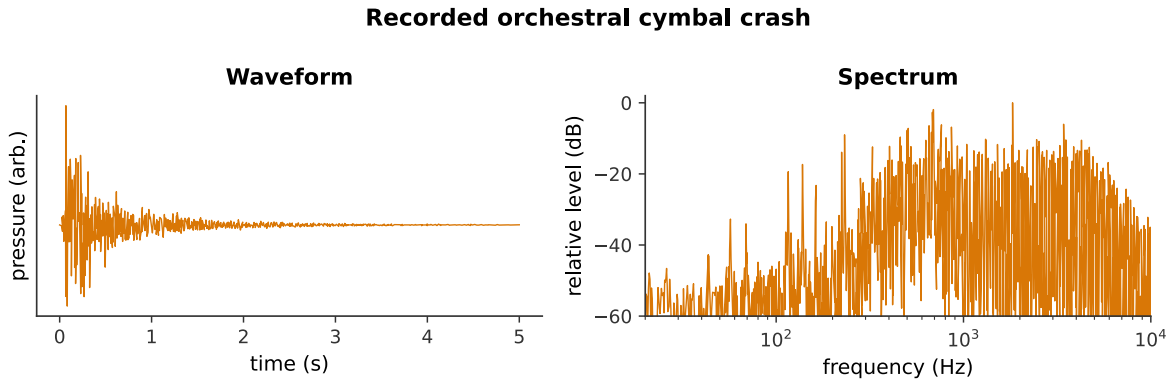


Figure 12.13: The real orchestral instrument. Its continuously filled spectrum makes clear why a list of a few ideal modes cannot capture a cymbal crash. CC0 recording from VSCO Community Edition. *What you hear:*A real pair of orchestral clash cymbals produces a sharp metallic impact followed by a dense, shimmering decay with no stable pitch. Listen: Real Cymbal

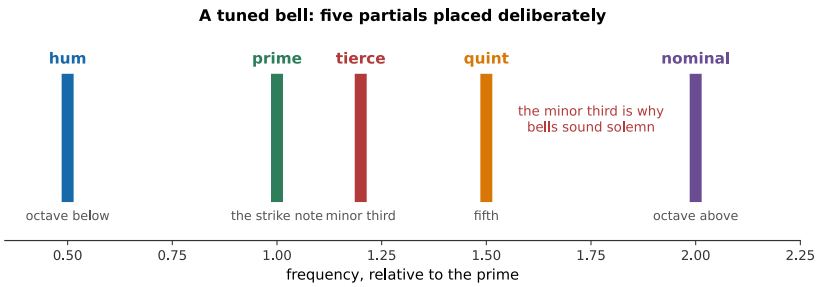


Figure 12.14: The five tuned partials of a church bell. Four are placed on consonant intervals; the **tierce** is deliberately a *minor* third, so bells sound solemn rather than bright, and a major-third bell, which founders can now make, sounds subtly wrong to most listeners.

partials, another missing fundamental. A bell's perceived pitch can be measured, and it does not correspond to any single measured mode.

The minor third is a choice. Bell founders tune by turning the bell on a lathe and removing metal from the inside at specific heights, each of which affects different modes by different amounts. The traditional tierce is minor, and Dutch founders in the 1980s worked out how to produce a major-third bell. It sounds, to ears raised on the traditional kind, slightly wrong, which is a nice demonstration that expectation is doing work alongside acoustics.

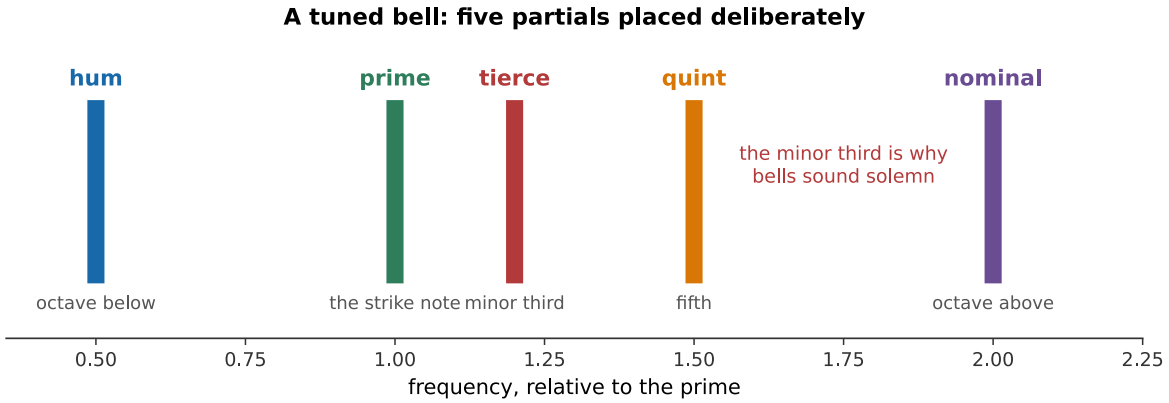


Figure 12.15: A bell with the five tuned partials and several above them. Listen to the *change* rather than the sound: the high partials die first, so the clangorous strike settles into the hum tone. That evolution is what makes a bell sound like a bell. *What you hear:*A bell strike: an initial clang containing many partials, resolving over several seconds into a steady low hum tone as the higher partials fade. Listen: Bell

12.5 Tuning the Untunable

Three Strategies

Percussion instruments with definite pitch are made so by one of three strategies. They are named together here because the whole chapter is instances of them.

Move the modes. Undercut a bar, turn a bell on a lathe, load a membrane with air. This is direct and it is what most tuned percussion does.

Suppress the modes that do not fit. Strike a timpano off-center to avoid the $(0, 1)$ piston mode. Support a bar at the fundamental's nodes.

Exploit the missing fundamental. Arrange for several partials to be harmonics of a note that is not there, and let the ear construct it. Timpani, tubular bells, and church bells all do this.

Pitch Where the Spectrum Is Inharmonic

"Pitch" for these instruments is not quite what it means for a violin.

For a harmonic instrument, the pitch is the fundamental, and the partials agree unanimously about what it is. For an inharmonic one, the partials disagree, and the perceived pitch is a **best compromise** the auditory system arrives at. It can therefore be ambiguous, some listeners hear a bell an octave away from others, and it can depend on how loud the sound is and what came before it.

This is not a defect of hearing. It is what happens when a mechanism evolved to group harmonically related partials is presented with partials that are not harmonically related, and the fact that it produces any answer at all is impressive.

The Steelpan as a Modern Solution

The steelpan, developed in Trinidad in the 1940s, is the newest acoustic instrument in this book and a genuine addition to the strategies above.

A section of oil drum is hammered into a concave dish, and distinct areas of the surface are shaped into **individual note regions**. Each region is a local resonator, tuned to a note, and the skill of the pan tuner is in shaping each so that its own second and third modes fall on the octave and twelfth above its fundamental.

So the steelpan turns one inharmonic plate into many nearly harmonic vibrators, and it does so by geometry alone, no undercutting, no air loading, no exploiting a missing fundamental. It is the best evidence in this chapter that tuned percussion is an engineering problem rather than a natural category, and that the problem is still being solved in new ways.

12.6 Summary

- **Two-dimensional and stiffness-controlled vibrators have inharmonic modes.** A circular membrane's are set by Bessel zeros ($1 : 1.59 : 2.14 : 2.30 : \ldots$); a free-free bar's by the roots of $\cos x \cosh x = 1$ ($1 : 2.76 : 5.40 : \ldots$).
- **Inharmonic partials give indefinite pitch**, because the pattern-matching mechanism of Chapter 8 finds no fundamental that explains them.
- **Undercutting a bar** removes material at the center, where the fundamental has an antinode and the second mode a node, lowering the first far more than the second. Xylophone bars are tuned to $3f_1$, marimba bars to $4f_1$.
- **Bars are supported at the fundamental's nodes**, about 22% from each end, so that holding them does not damp the note.
- **A timpano's kettle loads the modes unequally**, moving them into the near-harmonic set $1 : 1.5 : 2 : 2.5 : 3$. The perceived pitch is a **missing fundamental** an octave below the lowest of these.
- **Striking a timpano a quarter of the way in** suppresses the $(0, 1)$ piston mode, which is the loudest radiator and the one that does not fit the series.
- **Chladni figures** show nodal lines directly, and are still used to check the symmetry of violin plates.
- **A hard-struck cymbal is nonlinear**: energy cascades from low modes into high ones over a few tenths of a second, which is the audible "bloom".
- **A tuned bell has five named partials**, hum, prime, tierce, quint, nominal, placed deliberately, with the tierce a *minor* third by tradition. The strike note is a construction of the ear rather than a measured mode.
- **Tuned percussion uses three strategies**: move the modes, suppress the ones that do not fit, or exploit the missing fundamental. The steelpan adds a fourth by making many small resonators out of one plate.

12.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity. One question uses the chapter's Chladni-pattern figure as a visual prompt.

1. **Multiple choice.** Where should a free-free bar be supported to preserve its fundamental vibration: the center antinode, both ends, or the nodes about 22% from each end?

2. **Visual true or false.** In a Chladni pattern, sand collects along antinodes where the plate moves most.

3. **Drag the words.** Match the initial modal-frequency ratios $1 : 2$, $1 : 2.76$, $1 : 1.59$, and $1 : 1.5$ to an ideal string, free-free bar, circular membrane, and timpano.

4. **Fill in the blanks.** The perceived timpani pitch is a _____ fundamental one _____ below its lowest prominent partial.

5. **Mark the words.** Identify the five named bell partials: "The five named partials are hum, prime, tierce, quint, and nominal; the traditional tierce is a minor third."

12.8 Conceptual Questions

1. Explain why a one-dimensional string has harmonic modes and a two-dimensional membrane does not, without invoking Bessel functions by name.

2. A drum and a violin are struck and bowed respectively. Explain why one has a definite pitch and the other does not, in terms of what the ear is trying to do.

3. Explain why removing wood from the center of a marimba bar lowers the fundamental much more than the second mode.

4. Explain why a marimba bar is supported at about 22% of its length from each end rather than at the ends themselves.

5. A timpanist strikes about a quarter of the way in from the rim rather than at the center. Give two reasons.

6. The pitch a listener assigns to a timpano corresponds to no mode of the instrument. Explain how this can be.

7. A glockenspiel's bars are not undercut and yet it sounds pitched. Explain why it gets away with it.

8. Explain what is happening in a hard-struck cymbal that does not happen in a lightly struck one, and why this violates an assumption made throughout the book.

12.9 Problems

Exercise 12.1

(Straightforward) A circular membrane's lowest mode is at 180 Hz. Using the ratios 1, 1.59, 2.14, 2.30, 2.65: (a) Find the next four mode frequencies. (b) Is there any fundamental of which these are all harmonics?

Solution to Exercise 12.1

(a) Multiplying through:

$$286 \text{ Hz}, \quad 385 \text{ Hz}, \quad 414 \text{ Hz}, \quad 477 \text{ Hz}. \tag{12.5}$$

(b) No. For these to be harmonics of some f_0, every ratio between them would have to be a ratio of small whole numbers. But $286/180 = 1.59$, and 1.59 is not a ratio of small integers: the nearest is $8/5 = 1.60$, and even accepting that, 2.14 and 2.30 cannot be fitted to the same f_0 consistently.

Therefore there is no fundamental, the ear's pattern matcher returns nothing, and the membrane has no definite pitch.

Exercise 12.2

(Moderate) A marimba bar's plain modes are at 262, 722, and 1416 Hz. (a) Express the ratios. (b) After undercutting, the second mode is at 1048 Hz. What ratio is that? (c) What musical interval?

Solution to Exercise 12.2

(a) $722/262 = 2.756$ and $1416/262 = 5.405$.

(b) $1048/262 = 4.00$.

(c) A ratio of 4 is $1200 \log_2 4 = 2400$ cents, two octaves exactly.

Therefore the undercutting has moved the second mode from 2.76 to 4.00, turning a clashing partial into a consonant one. The shift is $1200 \log_2(4/2.756) = 645$ cents, over five semitones of retuning achieved with a gouge.

Exercise 12.3

(Moderate) A timpano's measured modes are 146, 219, 292, 356, 423 Hz. (a) Express them as ratios of the lowest. (b) Show that they are close to harmonics 2 to 6 of a missing fundamental. (c) What is that fundamental, and what note does the listener hear?

Solution to Exercise 12.3

(a) Dividing by 146: 1.00, 1.50, 2.00, 2.44, 2.90.

(b) Doubling: 2.00, 3.00, 4.00, 4.88, 5.79, close to $2, 3, 4, 5, 6$.

(c) The fundamental is $146/2 = 73$ Hz.

Therefore the listener hears about 73 Hz, D2, which is an octave below the lowest mode present and which the instrument radiates essentially nothing at. The fourth and fifth of these are a little flat of exact harmonics, so a timpano's pitch, while definite, is less crisply defined than a string's.

Exercise 12.4

(Moderate) A membrane's fundamental is $f_{01} = (\alpha_{01}/2\pi a)\sqrt{T/\sigma}$ with $\alpha_{01} = 2.405$. A timpano head has radius 0.33 m and mass per unit area 0.26 kg/m^2. (a) What tension per unit length gives a fundamental of 146 Hz? (b) By what factor must the tension change to raise the pitch a whole tone?

Solution to Exercise 12.4

(a) Rearranging:

$$T = \sigma \left(\frac{2\pi a f}{\alpha_{01}} \right)^2 = (0.26\ \text{kg/m}^2) \left(\frac{2\pi (0.33\ \text{m})(146\ \text{Hz})}{2.405} \right)^2 . \tag{12.6}$$

The bracket is $302.7/2.405 = 125.9$ m/s, so

$$T = (0.26\ \text{kg/m}^2)(125.9\ \text{m/s})^2 = 4.12 \times 10^3\ \text{N/m}. \tag{12.7}$$

(b) Frequency goes as $\sqrt{T}$, and a whole tone is a factor $2^{2/12} = 1.1225$:

$$\frac{T'}{T} = 1.1225^2 = 1.26. \tag{12.8}$$

Therefore about 4.1 kN/m of tension, and a 26% increase for a whole tone: a timpano's pedal therefore has to move a considerable distance, and the instrument's range is limited to about a fifth.

Exercise 12.5

(Moderate) A xylophone bar's second mode is tuned to $3f_1$ and a marimba bar's to $4f_1$. If both have $f_1 = 440$ Hz: (a) Find each second mode. (b) Which sounds brighter, and why? (c) Which is a more consonant relationship?

Solution to Exercise 12.5

(a) Xylophone: 1320 Hz. Marimba: 1760 Hz.

(b) The **xylophone** sounds brighter, which is initially counterintuitive since its second mode is lower. The reason is relative strength: a partial at $3f_1$ is a twelfth above the fundamental and sits in a region of high ear sensitivity, whereas at $4f_1$ it is two octaves up and, for the same amplitude, contributes less to the sense of a hard, edgy tone. The xylophone's shorter, less deeply cut bars also decay faster, which concentrates the energy into a sharper attack.

(c) Both are exact harmonics and therefore consonant. 4:1 is the simpler ratio and blends more completely into the fundamental, so the marimba sounds warm and unified while the xylophone sounds like a fundamental with a distinct twelfth on top.

Exercise 12.6

(Moderate) A glockenspiel bar has $f_1 = 2093$ Hz and is not undercut. (a) Where is its second mode? (b) Why does the inharmonicity matter less than it would on a marimba? (c) Estimate how many cents the second mode is from the nearest harmonic.

Solution to Exercise 12.6

(a) $2093 \times 2.756 = 5768$ Hz.

(b) Chapter 8 showed that pitch perception degrades above about 5 kHz, where phase locking fails. At 5768 Hz the partial contributes brightness but the auditory system is not attempting to fit it into a harmonic pattern, so its failure to fit costs nothing.

(c) The nearest harmonic is the third, at $3 \times 2093 = 6279$ Hz:

$$1200 \log_2 \left(\frac{5768}{6279} \right) = -147\ \text{cents}. \tag{12.9}$$

Therefore the second mode is nearly a semitone and a half from the nearest harmonic: an enormous error by the standards of Chapter 9, and entirely inaudible as mistuning because of where it falls.

Exercise 12.7
(Straightforward) A church bell's partials are at 131, 262, 314, 393, and 524 Hz. (a) Identify each by name. (b) Express the tierce as an interval above the prime. (c) What would a major-third bell's tierce be?

Solution to Exercise 12.7
(a) Taking 262 Hz as the prime: 131 is the **hum** (an octave below), 262 the **prime**, 314 the **tierce**, 393 the **quint**, 524 the **nominal** (an octave above).

(b) $314/262 = 1.198$:

$$1200 \log_2(1.198) = 313 \text{ cents}, \tag{12.10}$$

which is a minor third (the just minor third 6:5 is 316 cents).

(c) A major third would be 5:4:

$$262 \times 1.25 = 328 \text{ Hz}. \tag{12.11}$$

Therefore this is a conventional minor-third bell, and a major-third bell would need its tierce moved up 14 Hz, about 73 cents, which is a great deal of metal to remove from exactly the right place.

Exercise 12.8
(Straightforward) A timpanist strikes a head at its exact center. (a) Which modes are excited? (b) Which are not? (c) Describe the resulting sound and explain why timpanists never do this.

Solution to Exercise 12.8
(a) Only the modes with **no nodal diameter**: $(0,1)$, $(0,2)$, $(0,3)$, …These are the circularly symmetric ones, which move at the center.

(b) Every mode with $m \geq 1$, $(1,1)$, $(2,1)$, $(3,1)$, and so on, because all of them have a nodal diameter through the center.

(c) The $(m,1)$ family is precisely the near-harmonic set that gives the timpano its pitch, so striking at the center removes it entirely. What remains is the $(0,n)$ family, whose ratios are $1 : 2.30 : 3.60$, thoroughly inharmonic, plus the $(0,1)$ piston mode, which radiates very loudly and decays fast.

Therefore a center strike produces a loud, dull, pitchless thud. Timpanists strike about a quarter of the way in precisely to avoid it, which puts the mallet near a node of $(0,1)$ and at a strong antinode of $(1,1)$.

Exercise 12.9
(Challenging) Show that supporting a free-free bar at 22.4% of its length from each end places the supports at nodes of the fundamental, and explain what would happen if they were placed at the center instead.

Solution to Exercise 12.9

The fundamental transverse mode of a free-free bar has **two** nodes, symmetrically placed. Solving the bar equation places them at approximately $0.2242L$ and $0.7758L$: that is, at 22.4% from each end.

A support at a node touches the bar where the fundamental is not moving. It therefore removes no energy from that mode, and the note rings.

Supporting at the center would put the support at an **antinode** of the fundamental: the point of maximum motion. The support would absorb energy at the greatest possible rate, and the fundamental would be heavily damped. What would survive is the second mode, which has a node at the center.

Therefore a centrally supported bar would sound its second mode, weakly and briefly, at 2.76 times the intended pitch. The node placement is not a convenience; it is what makes the instrument work.

Exercise 12.10

(Moderate) A cymbal is struck softly and then hard. (a) Describe the difference in how the spectrum evolves. (b) Which assumption made throughout this book fails in the second case? (c) Name one other place in the book where the same assumption fails.

Solution to Exercise 12.10

(a) Struck softly, the cymbal behaves linearly: the modes excited by the strike ring and decay independently, and the spectrum only ever gets thinner. Struck hard, energy transfers *between* modes after the strike, so the spectrum broadens over a few tenths of a second before it decays: the sound blooms.

(b) **Superposition** (Chapter 3), which requires the restoring force to be proportional to displacement. At large deflection a plate's restoring force is not, and the modes cease to be independent.

(c) A brass instrument played very loudly (Chapter 11): the wave steepens as it travels down the bore, because compressions travel faster than rarefactions, adding brightness to a fortissimo that is not present at lower dynamics.

Therefore nonlinearity is rare in this book but not absent, and in both cases it is audible and musically useful rather than a defect.

Exercise 12.11

(Moderate) A steelpan note region has a fundamental of 392 Hz. Its tuner aims to place the second and third modes on the octave and the twelfth. (a) What frequencies are those? (b) Explain why this makes the note pitched. (c) Contrast the strategy with the timpani's.

Solution to Exercise 12.11

(a) Octave: 784 Hz. Twelfth: 1176 Hz.

(b) The three modes are then at $1 : 2 : 3$ times 392 Hz: an exact harmonic series. Chapter 8's pattern matcher finds 392 Hz immediately and unambiguously, and the note has a pitch as clear as a string's.

(c) The timpani places its modes at $2 : 3 : 4 : 5 : 6$ of a fundamental that is **not present**, and relies on the ear to construct it. The steelpan places its modes at $1 : 2 : 3$ with the fundamental **present and strong**. The steelpan's is the more direct solution and gives a clearer pitch; the timpani's is forced on it by the fact that a membrane's lowest mode cannot be moved into the series.

Therefore the two instruments solve the same problem by different strategies from §12.5, and the difference in clarity of pitch between them is audible.

Exercise 12.12

(Moderate) A bar's mode frequencies scale as $f \propto h/L^2$, where h is thickness and L length. A marimba bar sounds 262 Hz. (a) What length would sound 131 Hz at the same thickness? (b) Why is this a problem, and what do makers do instead?

Solution to Exercise 12.12

(a) Since $f \propto 1/L^2$, halving the frequency requires

$$L' = L\sqrt{2} = 1.414\,L. \tag{12.12}$$

(b) That is only a 41% increase in length for a whole octave, which sounds convenient until you notice the consequence for the instrument's **range**. Five octaves would need a length ratio of $(\sqrt{2})^5 = 5.7$, so if the top bar is 12 cm the bottom is 68 cm, which is manageable. The real problem is the other direction: to keep the bars a reasonable width and the instrument playable, the low bars must also be made **thinner**, and a thin bar is fragile and radiates poorly.

Makers therefore combine the levers: low bars are longer, thinner, and more deeply undercut, and a resonator tube is tuned beneath each to make up the radiation the thin bar cannot provide on its own.

Therefore the scaling law explains the shape of a marimba directly, and the resonator tubes exist because the scaling alone would leave the bass inaudible.

Exercise 12.13

(Challenging) Explain why a gamelan ensemble sounds in tune to a listener familiar with it and out of tune to one who is not, using this chapter and Chapter 8 together.

Solution to Exercise 12.13

Gamelan instruments are metallophones and gongs: bars, plates and shells, whose partials are inharmonic in the ratios this chapter has been describing.

Chapter 8 showed that the consonance of an interval is not a property of its ratio but of whether the two tones' *partials* coincide or clash. For instruments with harmonic partials, the intervals that make partials coincide are the simple ratios, and the Western scale sits on them. For instruments with inharmonic partials, the intervals that make partials coincide are somewhere else, and the slendro and pelog scales sit on those.

So a gamelan is in tune *for its own instruments*. Played on a piano, its intervals would clash, because the piano's partials are harmonic and the coincidences would not occur. Played on gamelan instruments, the Western scale would clash for the same reason in reverse.

A listener's judgment follows what they are used to hearing, and both judgments are correct about different things: the unfamiliar listener is hearing a real acoustic mismatch between the intervals and their *expectations*, which were formed on harmonic instruments.

Therefore neither tradition is approximating the other. Each has chosen intervals that minimize roughness for the instruments it actually has, which is the same procedure producing different answers.

Exercise 12.14

(Moderate) A tubular bell's strong partials are at 436, 654, and 872 Hz. (a) Express them as ratios of the lowest. (b) Find the missing fundamental. (c) Does the tube have a mode at that frequency? (d) Comment on what this means for how the instrument is notated.

Solution to Exercise 12.14

(a) $1.00 : 1.50 : 2.00$.

(b) Multiplying by two gives $2 : 3 : 4$, so these are harmonics 2, 3 and 4 of

$$f_0 = \frac{436}{2} = 218 \text{ Hz}. \tag{12.13}$$

(c) No. A tube's modes are in the inharmonic ratios of a bar, and 218 Hz corresponds to none of them; the tube does not vibrate there and radiates essentially nothing there.

(d) The instrument is notated, played, and conducted at 218 Hz: a note that exists only in the listener's auditory system. A microphone at the bell would find no energy at the written pitch.

Therefore tubular bells are a case where the *musical* fact and the *acoustic* fact genuinely differ, and the musical fact is the one everyone acts on. It is the same situation as the brass pedal tone in Chapter 11 and the organ's resultant stop in Chapter 8.

Chapter 13

The Singing Voice

Learning Objectives

By the end of this chapter, you should be able to:

- Describe the voice as a source-filter system, and identify the anatomical structure serving as each.
- Explain how the vocal folds oscillate, and why the oscillation is self-sustaining rather than driven at a chosen frequency.
- Relate the fundamental frequency of phonation to vocal-fold length, mass, and tension, and explain the typical difference between adult voices.
- Describe the spectrum of the glottal source, and state approximately how its harmonic amplitudes fall with frequency.
- Define a formant, explain how the vocal tract produces formants, and calculate the formant frequencies of a uniform tube of the length of the adult vocal tract.
- Identify vowels from the frequencies of the first two formants, and read a vowel chart in the F_1-F_2 plane.
- Explain why the vowel of a sung note is harder to identify at high pitch, in terms of harmonic spacing relative to formant width.
- Describe the vocal registers and what changes physically between them.
- Explain the singer's formant, state approximately where it lies, and explain how it lets an unamplified soloist be heard over an orchestra.

Introduction

The voice is the oldest instrument and the only one everybody has. It is also the one that shows most clearly a model this book has been assembling since Chapter 4: the **source–filter model**.

The idea is that an instrument can be described as two independent stages. A **source** produces a raw signal with a rich spectrum, and a **filter** shapes that spectrum by emphasizing some frequencies and suppressing others. The two can be varied independently, and what a listener hears is the product.

In the voice, the separation is unusually clean. The **vocal folds** are the source: they chop the airflow into puffs, producing a harmonic series whose fundamental is the pitch. The **vocal tract**, the tube from the larynx to the lips, is the filter: its resonances, called **formants**, emphasize whichever harmonics fall near them.

The crucial point, and the one that makes speech possible at all, is that **the two are**

independent. You can change the pitch without changing the vowel, by tensing the folds; and you can change the vowel without changing the pitch, by moving the tongue and jaw. Singing is exactly the ability to do both at once, deliberately.

This chapter also explains one thing that ought to be impossible: how a single unamplified voice is heard over a hundred-piece orchestra. The answer is a trick of spectrum placement, and it is a good one.

13.1 The Voice as Source and Filter

Two Independent Stages

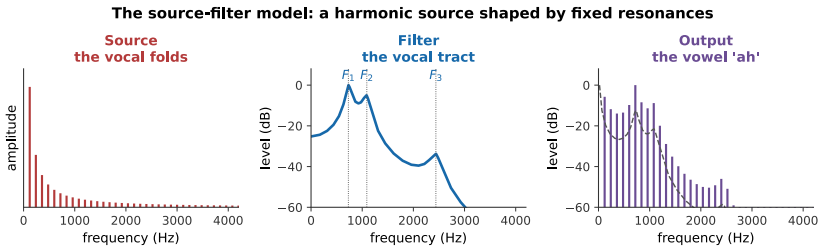


Figure 13.1: The source–filter model. The source supplies a harmonic series at the pitch being sung. The filter has fixed resonances. The output is their product: the same harmonics, with their amplitudes reshaped. Note that the filter is drawn in decibels, the third formant is some 35 dB below the first, and on a linear axis it would be invisible.

The model is an idealization, and the idealization is good. There *is* some interaction, the tract's pressure fluctuations feed back on the folds, and at extremes of range a singer can exploit this deliberately, but for most purposes the two stages can be treated as independent, and doing so explains a great deal.

Why the Model Works

The reason the separation holds so well is a difference of impedance, and it is the same argument as Chapter 11's reed.

The vocal folds present a high impedance to the tract: they are relatively massive and stiff, and the tract's pressure oscillations are not strong enough to push them around much. So the tract does not tell the folds what to do, and the source frequency is set by the folds alone.

This is the **opposite** of a wind instrument, where the bore controls the reed. Acoustically, then, the voice is more like a string instrument, source sets pitch, resonator shapes tone, than like the wind instrument it superficially resembles.

The Anatomy in Brief

Three parts, in the order air meets them.

The **lungs and diaphragm** supply a steady pressure: a few hundred pascals above atmospheric for normal speech, more for loud singing. This is the power supply, and the entire art of breath support is keeping it steady.

The **larynx** houses the vocal folds: two layered flaps of muscle and ligament, about 17–25 mm long in adult men and 12–17 mm in adult women, covered in mucous membrane. They can be brought together to close the airway or held apart to breathe.

The **vocal tract** runs from the folds to the lips, about 17 cm in an adult man. It includes the pharynx, the mouth, and, when the soft palate is lowered, the nasal cavity. Its shape can be changed continuously by the tongue, jaw, lips, and larynx height.

13.2 The Source: The Vocal Folds

Self-Sustained Oscillation

The folds are not vibrated by anything. They oscillate **themselves**, driven by a steady airflow, and the mechanism is a feedback loop with no oscillating input at all.

The cycle runs like this. Air pressure below the closed folds builds until it forces them apart. Air rushes through the gap, and, by the Bernoulli effect, the fast-moving air in the narrow gap is at *reduced* pressure, which sucks the folds back together. Their own elasticity helps. They close, the pressure builds again, and the cycle repeats.

The essential feature is that the folds do not move as a single stiff object. They have a layered structure, and the mucous membrane travels over the deeper tissue as a surface wave, so the lower edge opens before the upper edge and closes before it too. That phase difference between the opening and the closing is what allows net energy to be extracted from a steady airflow, without it, the work done pushing the folds open would exactly cancel the work recovered as they close, and the oscillation would die.

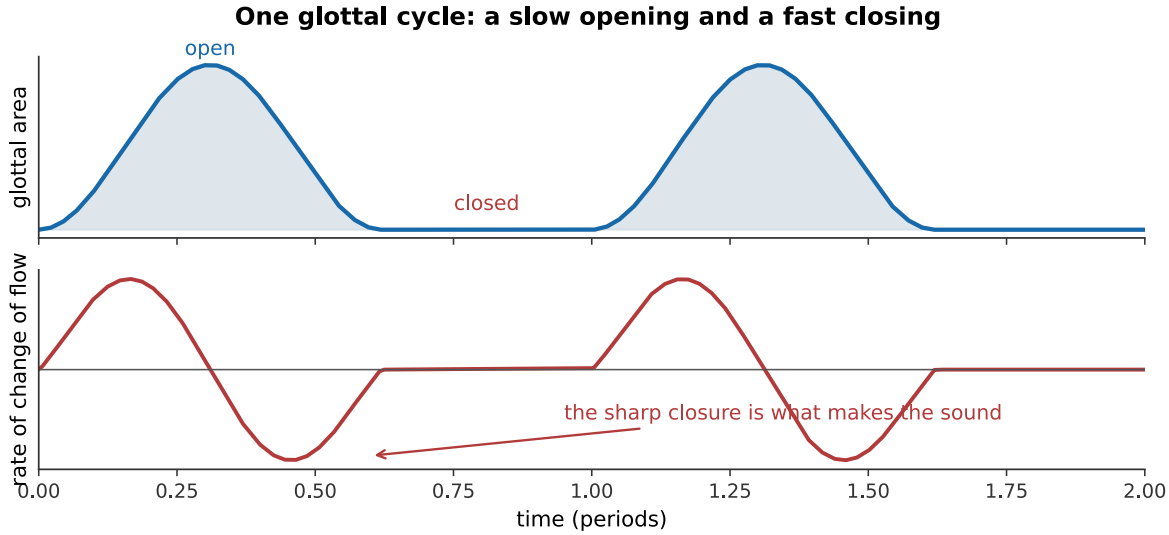


Figure 13.2: One glottal cycle. The opening is gradual and the closing is abrupt, and the folds stay shut for a substantial fraction of each cycle. **The sharp closure is what makes the sound**: it produces a rapid change in airflow, which is a broadband excitation of the tract, in the same way that a sharp-cornered waveform contains high harmonics (Chapter 5).

What Sets the Fundamental

The folds are a mass–spring system, so Chapter 1 applies: frequency rises with tension and falls with mass.

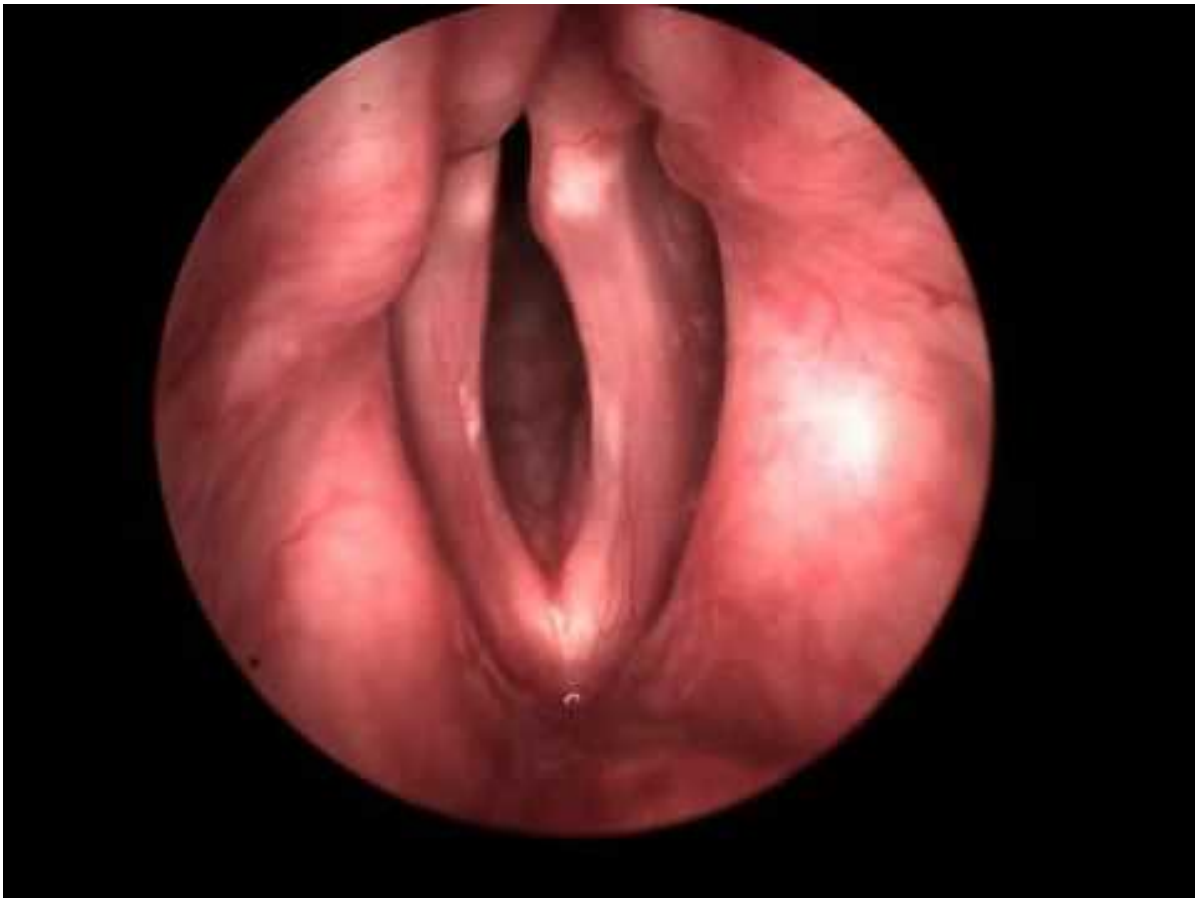


Figure 13.3: High-speed laryngeal imaging slows the vocal folds' oscillation enough to reveal the opening and closing cycle. The footage shows the vibrator that supplies the source in the source–filter model. Watch on YouTube: High-Speed Video of the Vocal Folds

Change	Effect on pitch	How a singer does it
Increase fold tension	Higher	Cricothyroid muscle tilts the thyroid cartilage
Increase effective mass	Lower	Thicker, looser folds; chest voice
Increase subglottal pressure	Slightly higher	More breath pressure

That last row is a nuisance for singers: getting louder tends to sharpen the pitch, so a singer must actively compensate. It is also why breath support and intonation are taught together.

The anatomical difference between voice types is straightforward. Adult male folds are longer and more massive, hence roughly 85–180 Hz in speech, while adult female folds are shorter and lighter, at roughly 165–255 Hz. The ratio is close to a factor of two, so men and women speaking naturally are about an octave apart.

The Spectrum of the Glottal Pulse

The glottal waveform is not a sine. It is a series of pulses, and Chapter 5 guarantees that a periodic pulse train has a full harmonic series.

The amplitudes fall at roughly 12 dB per octave: that is, as $1/n^2$. This is the **source spectrum**, and it is important that it is smooth: it has no peaks of its own, so every peak in the output belongs to the filter.

A singer *can* change the source spectrum. Pressed, forceful phonation with a sharper closure produces a brighter source; breathy phonation with incomplete closure produces a weaker, noisier one with a steeper roll-off. These are real timbral controls, but they are secondary to what the tract does.

13.3 The Filter: The Vocal Tract

A Tube, Stopped at One End

To a first approximation the vocal tract is a uniform tube about 17 cm long, closed at the glottis and open at the lips: a **stopped pipe**, exactly as in Chapter 3:

$$f_n = n\,\frac{v}{4L}, \qquad n = 1, 3, 5, \ldots \tag{13.1}$$

With $L = 0.17$ m and $v = 350$ m/s (warm, moist air):

$$f_1 = \frac{350\ \text{m/s}}{4(0.17\ \text{m})} = 515\ \text{Hz}, \qquad f_3 = 1545\ \text{Hz}, \qquad f_5 = 2575\ \text{Hz}. \tag{13.2}$$

Roughly 500, 1500, 2500 Hz, and those are, to a remarkable degree, the measured formant frequencies of a neutral vowel. A crude stopped-pipe model gets the human vocal tract approximately right.

Formants

A **formant** is a resonance of the vocal tract. It is a property of the *tube*, not of the pitch being sung, and this is the single most important fact in the chapter.

Formants are numbered upward: F_1, F_2, F_3. Each has a bandwidth of 50–150 Hz, which is a Q of order ten, deliberately low, so that the tract responds usefully over a wide range of pitches rather than only at isolated frequencies (Chapter 4).

Formants do not move when the pitch does. Sing "ah" on a low note and then on a high one: the fundamental changes, every harmonic changes, and the formants stay exactly where they were, because the shape of your throat has not changed. That is why the vowel is still recognizable.

Worked example: what leaves the mouth

A baritone sings $F_0 = 110$ Hz. His glottal source spectrum falls at 12 dB per octave, that is, amplitude $\propto 1/n^2$. His first formant sits at 550 Hz, exactly on the 5th harmonic, with a peak gain of 15 dB over the surrounding response. Find the level of the 5th harmonic relative to the fundamental (a) in the source alone, (b) after the filter.

(a) The 5th harmonic's source amplitude, relative to the fundamental's, is $1/5^2 = 1/25$:

$$20 \log_{10}\left(\frac{1}{25}\right) = -28 \text{ dB}. \tag{13.3}$$

(b) The formant sits exactly on this harmonic, adding its full 15 dB of peak gain:

$$-28 + 15 = -13 \text{ dB}. \tag{13.4}$$

Therefore, on its way from the folds to the lips, the 5th harmonic starts 28 dB below the fundamental and leaves only 13 dB below it: the formant has clawed back more than half its disadvantage. This is the source–filter model doing exactly what its name promises, source and filter multiply, which in decibels means their contributions simply add, and it is why a formant's placement matters as much as a source's raw output.

Shaping the Tube

A singer changes formants by changing the tract's shape, and the controls map onto the first two formants in a usefully simple way.

F_1 tracks how open the tract is. Constricting the tract near the front raises F_1; opening the jaw raises it. Roughly, F_1 says how open your mouth is.

F_2 tracks where the tongue is. A tongue hump far forward raises F_2; a tongue pulled back lowers it.

F_3 and above depend on finer details, including the space under the tongue and the position of the larynx, and §13.5 shows that F_3 is where trained singing does something distinctive.

13.4 Vowels, Registers, and Range

Vowels in the Formant Plane

Because F_1 and F_2 are controlled by two roughly independent gestures, vowels occupy a two-dimensional space.

The identification is robust. Two speakers with very different vocal tract lengths produce quite different absolute formant frequencies for the same vowel, yet listeners have no difficulty: the auditory system evidently normalizes by the speaker rather than reading absolute values.



Figure 13.4: Real-time MRI makes the filter half of the voice visible. As Tyley Ross changes vowels and pitch, the tongue, jaw, lips, and vocal tract reshape the resonant cavities that determine the formants. Watch on YouTube: Singing in the MRI: Making the Voice Visible

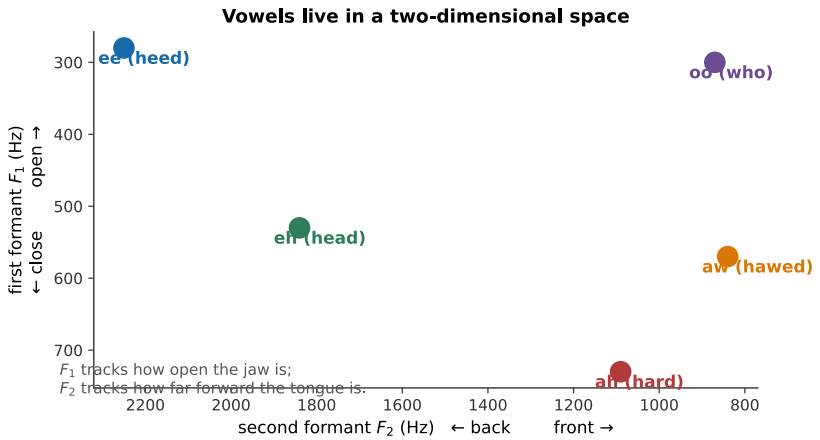


Figure 13.5: Vowels plotted by their first two formants. With both axes reversed, the plot reproduces the vowel quadrilateral that phoneticians drew from articulation long before formants could be measured, which is a nice confirmation that the acoustic and articulatory descriptions are the same thing.

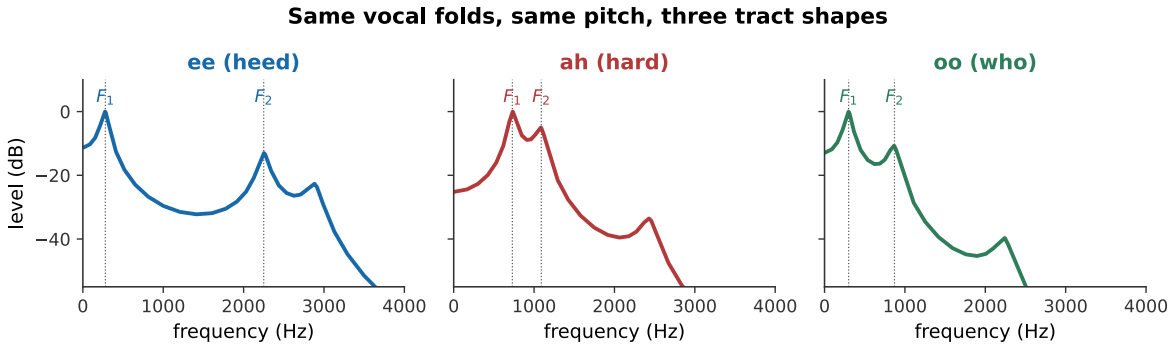


Figure 13.6: Three vowels at 130 Hz. The source is identical in all three, the same harmonic series at the same pitch. Only the filter differs, and it is enough to produce three recognizably different speech sounds. *What you hear:*Three sung vowels at the same pitch. They are clearly different vowels and clearly the same note, sung by the same voice. Listen: ee, ah, oo

Formant Tuning at High Pitch

Now a problem that is specific to singing, and it is the reason opera has the reputation it does for incomprehensibility.

The tract's formants are sampled by the harmonics of whatever is being sung. At a low pitch the harmonics are closely spaced, so there is always one near any formant, and the formant's peak is well represented in the output. At a high pitch they are widely spaced, and a formant can fall in the gap between two harmonics, in which case there is *nothing in the output* to tell the listener the formant is there.

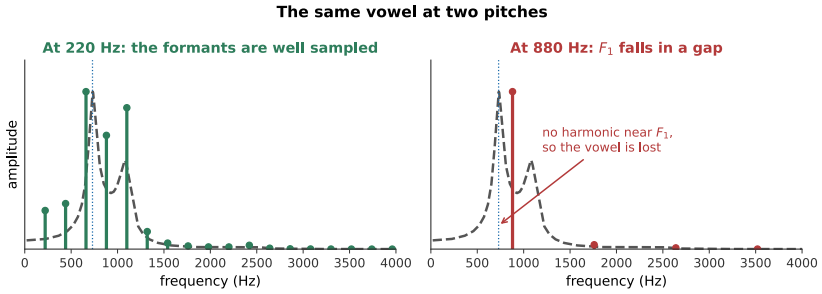


Figure 13.7: The same vowel, sung at 220 Hz and 880 Hz. At the low pitch the harmonics sample the filter closely and the formant peaks are faithfully rendered. At the high pitch the first formant falls between harmonics, and its peak simply does not appear in the output.

Sopranos deal with this by **formant tuning**: they modify the vowel, opening the jaw to raise F_1 until it coincides with a harmonic, usually the fundamental itself. The note becomes much louder, because a harmonic is now sitting on a resonance peak, and the vowel becomes whatever F_1 placement requires. Above about C6 a soprano is essentially singing one vowel whatever the text says, and this is a physical necessity rather than a failure of technique.

Registers and the Passaggio

Registers are distinct modes of vocal fold vibration, and they are physically real rather than metaphorical.

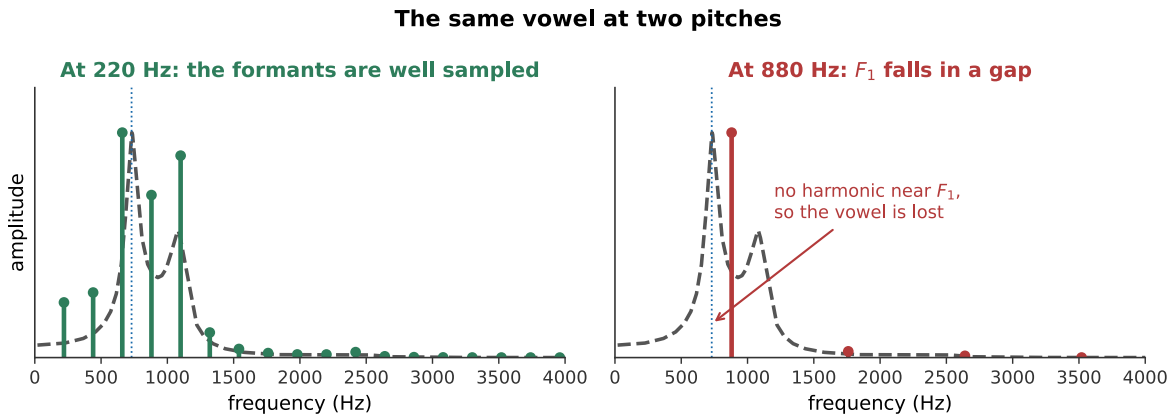


Figure 13.8: The vowel at a low and a high pitch, with an identical filter. The vowel identity degrades at the top, and no amount of effort by the singer can put information into a frequency region where there is no harmonic to carry it. *What you hear:*The same vowel at two pitches two octaves apart. The low one is clearly ”ah”; the high one is much less definite, and could be almost any open vowel. Listen: ”ah” at 196 Hz, ”ah” at 784 Hz

In **chest voice** (or *modal* register) the folds vibrate over their full depth, closing completely on each cycle. The tone is strong and rich in harmonics.

In **head voice** (or *falsetto*) only the thin upper edges vibrate, and the folds may not close completely. The tone is weaker and much poorer in harmonics.

Between them lies the ***passaggio***, where the singer must change from one mode to the other. Left to itself the transition is abrupt and audible, the “break”, and a great deal of vocal training consists of learning to blend across it by adjusting fold tension and tract shape together so that the change is gradual.

The **whistle register**, available to some sopranos above about C6, involves only a small portion of the folds and produces a nearly pure tone with very few harmonics, which is exactly why it sounds flute-like and why the words are entirely lost.

13.5 Being Heard

The Singer’s Formant

Here is the problem. An orchestra playing fortissimo produces around 100 dB in the hall. A single unamplified human voice produces perhaps 90 dB at a meter, and less by the time it reaches the audience. Adding one voice to a hundred instruments should be inaudible, by Chapter 7’s addition rule.

Yet an opera singer is heard clearly. The trick is not loudness but **spectral placement**.

The **singer’s formant** is a strong peak around 2.8–3.2 kHz, and it is produced by clustering F_3, F_4, and F_5 together. The mechanism is lowering the larynx, which widens the pharynx just above it and creates a small, separately resonant cavity; that cavity’s resonance coincides with the upper formants and reinforces them.

Its effect is large: 20 dB or more of extra output in that band.

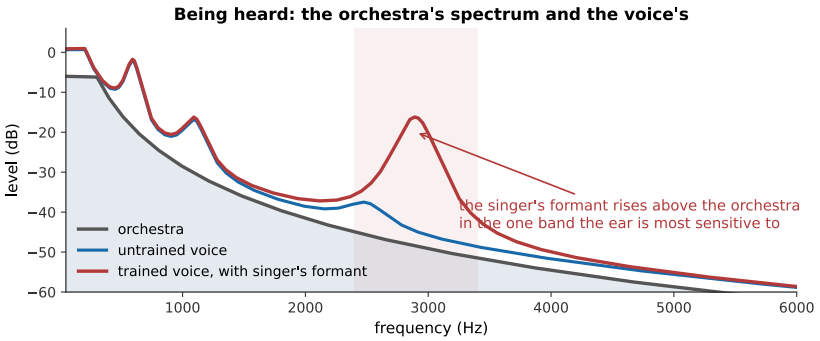


Figure 13.9: Why a trained soloist is heard. An orchestra's spectrum is concentrated below about 500 Hz and falls away steeply. A trained voice puts a strong resonance near 2.9 kHz, precisely where the orchestra has little energy and where the ear is most sensitive (Chapter 7).

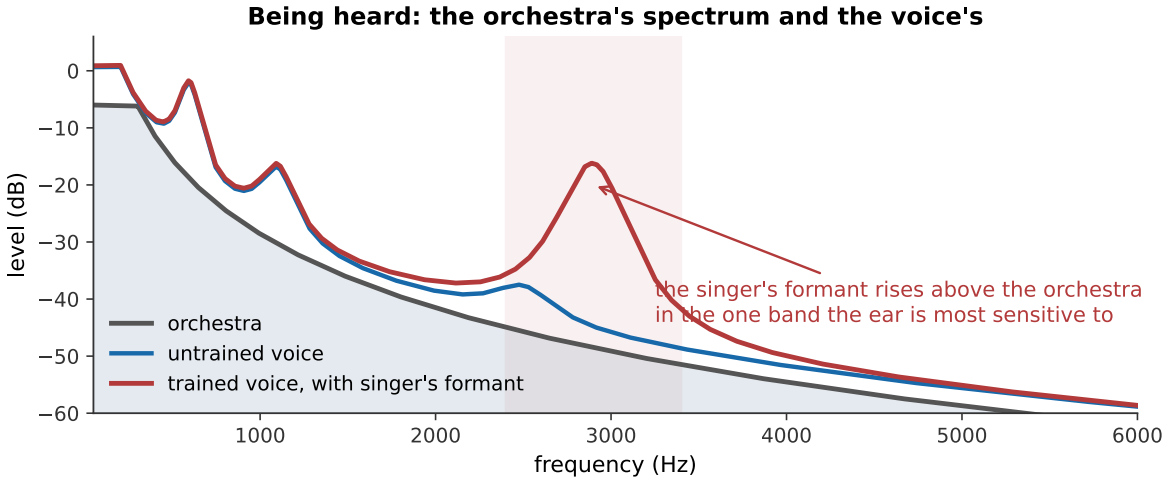


Figure 13.10: A voice mixed with a low-biased orchestral background, with and without the singer's formant. The overall level of the voice is the same in both. What has changed is *where* its energy sits. *What you hear:*The same voice against the same orchestral background twice. In the first it is buried and hard to follow; in the second it cuts through clearly, without being obviously louder overall. Listen: Without singer's formant, With singer's formant

Competing With an Orchestra

Three features of the solution are worth drawing together, because it is a genuinely clever piece of physics that nobody designed.

It exploits a **gap in the competition**. The orchestra's energy is concentrated low, so the 3 kHz region is comparatively empty.

It exploits the **ear's sensitivity peak**. Chapter 6 showed the ear canal resonates near 3.4 kHz, and Chapter 7 showed that hearing is most sensitive there. A decibel spent in that band buys more loudness than a decibel spent anywhere else.

It exploits **masking asymmetry**. Chapter 7 showed that masking spreads upward far more readily than downward, so low-frequency orchestral energy masks the voice's low harmonics severely and its 3 kHz region much less.

The same logic appears elsewhere. A violin's **bridge hill** is at 3 kHz (Chapter 10), and a trumpet's bell radiates most efficiently in the same region (Chapter 11). Several unrelated instrument traditions converged on the same band, because that is where the ear is listening.

Vibrato, Projection, and the Limits of the Voice

Vibrato is a periodic modulation of pitch, typically 5–7 Hz and ±50 to ±100 cents. It arises partly from a natural oscillation in the laryngeal muscles and partly from training.

Two things it does. It makes a voice **stand out** against a steady background, because the auditory system groups components that change together and separates those that do not: a moving signal against a static one is easier to follow. And because the harmonics sweep back and forth across the formant peaks, vibrato *samples* the filter over a range rather than at a point, which fills in the missing information of §13.4 and makes the vowel more identifiable at high pitch than a steady tone would be.

Finally, the limits. The voice is a small instrument. A trained singer produces perhaps 1 mW of acoustic power at full voice, comparable to a single violin, and against sixty of them. Its range is about two and a half octaves for most singers and rarely more than three. Its dynamic range is far narrower than an orchestra's.

Everything in this chapter is, in one way or another, a response to those constraints. The voice cannot win on power, so it wins on placement.

13.6 Summary

- **The voice is a source–filter system.** The vocal folds are the source, producing a harmonic series at the pitch; the vocal tract is the filter, whose resonances, **formants**, shape it. The two are essentially independent, so pitch and vowel can vary separately.
- **The folds oscillate on their own**, driven by steady airflow. The layered structure makes the upper and lower edges move out of phase, which is what allows net energy extraction from a steady flow.
- **The sharp closure makes the sound.** A rapid cut-off of airflow is a broadband excitation, and the glottal source spectrum falls smoothly at about 12 dB per octave with no peaks of its own.
- **Fundamental frequency rises with fold tension and falls with fold mass**, and rises slightly with breath pressure; so loud singing tends to go sharp.
- **The tract is approximately a 17 cm stopped pipe**, giving formants near 500, 1500, 2500 Hz. F_1 tracks jaw opening; F_2 tracks tongue position; vowels occupy the F_1–F_2

plane.

- **Formants do not move when the pitch does**, so a vowel survives across a singer's range, up to a point.
- **At high pitch the harmonics are too widely spaced to sample the formants.** Sopranos respond with **formant tuning**, opening the jaw to put F_1 on a harmonic, which gains loudness and loses vowel identity.
- **Registers are distinct modes of fold vibration**: chest voice uses the full fold depth, falsetto only the edges, and the *passaggio* between them is what training smooths.
- **The singer's formant**, near 2.9 kHz, is produced by clustering F_3 to F_5 via a lowered larynx. It exploits a gap in the orchestra's spectrum, the ear's sensitivity peak, and the asymmetry of masking, and it is worth 20 dB where it matters.
- **Vibrato** helps the ear separate the voice from a steady background, and sweeps the harmonics across the formants, restoring some of the information lost at high pitch.

13.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple response.** Why does the singer's formant near 2.9 kHz improve projection? Consider the orchestral spectrum, ear sensitivity, masking, and the vocal fundamental.

2. **True or false.** In the source–filter model, the vocal folds can change pitch while the vocal tract largely preserves the vowel filter.

3. **Drag the words.** Match vocal folds, vocal tract, jaw opening, and tongue position to the harmonic source, formant filtering, F_1, and F_2.

4. **Fill in the blanks.** A roughly _____ cm tract has formants near 500, _____, and 2500 Hz.

5. **Mark the words.** Identify two vocal registers: "Chest voice uses more of the fold depth, while falsetto uses mainly the edges; the passaggio is the transition between them."

13.8 Conceptual Questions

1. Explain what it means to say the voice is a source–filter system, and give one observation that shows the two stages are independent.

2. Explain why the vocal tract controls the *timbre* of the voice but not its pitch, and contrast this with a clarinet.

3. Explain why the abrupt closure of the vocal folds, rather than their opening, is what produces most of the sound.

4. A singer gets louder and the pitch drifts sharp. Explain the mechanism.

5. Explain why a vowel remains recognizable when the same person sings it an octave higher, and why this eventually fails.

6. A soprano singing above C6 is hard to understand whatever the language. Explain why this is a physical limitation rather than a matter of diction.

7. Explain how a single unamplified voice is heard over an orchestra, naming the three separate effects that make it work.

8. Explain two distinct benefits of vibrato, one about separating the voice from the background and one about vowel identity.

13.9 Problems

Exercise 13.1
(Straightforward) Model the vocal tract as a stopped pipe of length 17.5 cm, with $v = 350$ m/s in warm moist air. (a) Find the first three formant frequencies. (b) Compare with the measured neutral-vowel values of about 500, 1500, 2500 Hz.

Solution to Exercise 13.1
(a) A stopped pipe gives odd harmonics of $v/4L$:

$$f_1 = \frac{350 \text{ m/s}}{4(0.175 \text{ m})} = 500 \text{ Hz}, \quad f_3 = 1500 \text{ Hz}, \quad f_5 = 2500 \text{ Hz}. \tag{13.5}$$

(b) The agreement is essentially exact.

Therefore a uniform tube of the right length predicts the neutral vowel's formants almost perfectly. Departures from this pattern are what produce the *other* vowels: a vowel is a deviation from the neutral tube.

Exercise 13.2
(Moderate) An adult male tract is 17.5 cm and an adult female tract about 15 cm. (a) Find the neutral formants for each. (b) Express the difference as a musical interval. (c) How do listeners cope?

Solution to Exercise 13.2
(a) Male: 500, 1500, 2500 Hz. Female, with $L = 0.15$ m:

$$f_1 = \frac{350 \text{ m/s}}{4(0.15 \text{ m})} = 583 \text{ Hz}, \quad f_3 = 1750 \text{ Hz}, \quad f_5 = 2917 \text{ Hz}. \tag{13.6}$$

(b) The ratio is $583/500 = 1.167$:

$$1200 \log_2(1.167) = 267 \text{ cents}, \tag{13.7}$$

between a whole tone and a minor third.

(c) The auditory system **normalizes** for the speaker rather than reading absolute formant values. It uses the pattern of formants relative to one another, and cues from the speaker's fundamental and from surrounding speech, to calibrate. A child, a woman, and a man can all say "ah" and be understood as saying the same thing despite formants differing by a factor of 1.3.

Exercise 13.3
(Moderate) A bass sings at $F_0 = 98$ Hz and a soprano at $F_0 = 880$ Hz, both with $F_1 = 700$ Hz. (a) Which harmonic is closest to F_1 in each case? (b) How far is it, in hertz? (c) Comment on vowel clarity.

Solution to Exercise 13.3
(a) Bass: $700/98 = 7.14$, so the 7th harmonic at 686 Hz. Soprano: $700/880 = 0.80$, so the fundamental at 880 Hz is the nearest, there is nothing below it.

(b) Bass: $700 - 686 = 14$ Hz away. Soprano: $880 - 700 = 180$ Hz away.

(c) The bass has a harmonic essentially sitting on the formant, so the formant's peak appears faithfully in the output and the vowel is clear. The soprano's nearest harmonic is 180 Hz above the formant, well outside its 100 Hz bandwidth, so the formant contributes almost nothing and the vowel is poorly defined.

Therefore this is the formant-sampling problem of §13.4 in numbers, and it is why sopranos resort to formant tuning.

Exercise 13.4
(Moderate) A soprano at 880 Hz raises F_1 from 700 Hz to 880 Hz by opening her jaw. (a) What does this do to the output level of the fundamental? Assume the formant peak is 15 dB above the surrounding response. (b) What does it do to the vowel?

Solution to Exercise 13.4
(a) The fundamental moves from the skirt of the resonance onto its peak, gaining most of the 15 dB. In intensity terms:

$$10^{15/10} = 32, \tag{13.8}$$

a thirty-twofold increase in the fundamental's radiated intensity: an enormous gain for a change in jaw position.

(b) F_1 has been moved from 700 Hz to 880 Hz. Consulting the vowel chart, raising F_1 moves the vowel toward a more open one, "ah" territory and beyond, whatever the text calls for.

Therefore the singer trades intelligibility for power, and above a certain pitch she has no real choice: the alternative is to be inaudible. Sung text at the top of the soprano range is conventionally supported by surtitles for that reason.

Exercise 13.5
(Moderate) The singer's formant provides 20 dB of extra output near 3 kHz. (a) By what factor in intensity? (b) An orchestra produces 95 dB in the hall and an untrained voice 72 dB in the 3 kHz band. What does the trained voice produce there? (c) Comment.

Solution to Exercise 13.5
(a) $10^{20/10} = 100$.

(b) $72 + 20 = 92$ dB in that band.

(c) The orchestra's 95 dB is its *total* level across the spectrum; in the narrow 3 kHz band it produces far less, perhaps 70 dB. The trained voice's 92 dB in that band therefore **exceeds** the orchestra's contribution there by some 22 dB, even though the voice's total output is 20 dB below the orchestra's.

Therefore the voice loses the broadband contest decisively and wins the 3 kHz contest decisively, and the second is the one that determines whether it is heard.

Exercise 13.6
(Moderate) Vocal folds are modeled as a mass on a spring. A singer raises the pitch from 220 Hz to 440 Hz. (a) If the effective mass is unchanged, by what factor does the tension rise? (b) In practice the folds also thin as they stretch. How does that affect the answer?

Solution to Exercise 13.6
(a) $f \propto \sqrt{k}$, so doubling the frequency needs

$$\frac{k'}{k} = 2^2 = 4. \tag{13.9}$$

(b) Stretching the folds makes them thinner, which **reduces** the effective vibrating mass. Since $f \propto \sqrt{k/m}$, a falling m contributes to the rise in f, so less than a fourfold increase in stiffness is needed.

Therefore the naive answer of four times overestimates the tension required. This is one reason the vocal folds can cover two and a half octaves while a guitar string covers far less by tension alone: the folds change both levers at once.

Exercise 13.7
(Moderate) A singer's vibrato is 5.5 Hz at a depth of ±70 cents on a 440 Hz note. (a) What frequency range does the note sweep? (b) F_1 is at 700 Hz. Which harmonic sweeps across it, and by how much?

Solution to Exercise 13.7
(a) ±70 cents is a ratio of $2^{\pm 70/1200} = 1.0412$ and 0.9604:

$$440 \times 0.9604 = 422.6 \text{ Hz} \quad \text{to} \quad 440 \times 1.0412 = 458.1 \text{ Hz}. \tag{13.10}$$

(b) The nearest harmonic to 700 Hz is the 2nd, at 880 Hz, too high. Actually $700/440 = 1.59$, so no harmonic is near F_1 at all; the fundamental at 440 Hz is the closest below.

The fundamental sweeps from 422.6 to 458.1 Hz, a range of 35.5 Hz. Since F_1's bandwidth is around 100 Hz, this sweep moves the fundamental appreciably across the lower skirt of the resonance.

Therefore vibrato causes the output amplitude of each harmonic to fluctuate as it moves across the filter's slopes. Vibrato therefore produces amplitude modulation as well as frequency modulation, and makes a voice easier to pick out of a texture.

Exercise 13.8
(Moderate) Explain, using Chapter 7's masking asymmetry, why an orchestra masks a voice's low harmonics more effectively than its high ones.

Solution to Exercise 13.8
Chapter 7 established that masking spreads **upward** in frequency far more than downward, because a low-frequency traveling wave must traverse the whole basilar membrane to reach its own place and excites the high-frequency region on the way (Chapter 6).

The orchestra's energy is concentrated below about 500 Hz. That energy therefore masks:

- the voice's fundamental and low harmonics, which lie in the same region, strongly, since masker and masked coincide;
- the voice's mid harmonics, at 1–2 kHz, appreciably, by upward spread;

- the voice's 3 kHz region, much less, because the upward spread has weakened considerably by two and a half octaves above the masker.

Therefore the region where the voice is least masked is the region where the singer's formant puts its energy, and it is also where the ear is most sensitive. The three effects reinforce each other, and the solution works as well as it does.

Exercise 13.9

(Challenging) A trained singer produces about 1 mW of acoustic power. (a) What sound level does this give at 10 m outdoors? (b) An orchestra produces about 1 W. What level at the same distance? (c) Reconcile with the fact that soloists are heard.

Solution to Exercise 13.9

(a) $I = P/4\pi r^2 = (10^{-3}\ \text{W})/(4\pi \times 100\ \text{m}^2) = 7.96 \times 10^{-7}\ \text{W/m}^2$:

$$L = 10 \log_{10}\left(\frac{7.96 \times 10^{-7}}{10^{-12}}\right) = 59\ \text{dB}. \tag{13.11}$$

(b) A thousand times the power is 30 dB more: 89 dB.

(c) The voice is 30 dB below the orchestra **overall**, which by Chapter 7's addition rule should make it entirely inaudible: a sound 30 dB down adds nothing.

It is heard because the comparison is not made broadband. In the narrow 3 kHz band where the singer's formant concentrates the voice's output and where the orchestra has little, the voice is *above* the orchestra. The ear analyzes by critical band (Chapter 7), so what matters is the signal-to-masker ratio within a band, not the total.

Therefore the solution is to lose the overall contest and win a local one, and the ear's band-by-band analysis is what makes that a winning strategy.

Exercise 13.10

(Moderate) A tenor lowers his larynx, shortening the effective tract above the larynx from 17.5 cm to 16.0 cm while creating a separate cavity below. (a) What happens to the main formants? (b) The new cavity resonates near 2.9 kHz. Estimate its dimensions, treating it as a stopped pipe.

Solution to Exercise 13.10

(a) Shortening the tract raises the formants in proportion to $1/L$:

$$\frac{17.5}{16.0} = 1.094, \tag{13.12}$$

so F_1 moves from 500 to 547 Hz and F_2 from 1500 to 1641 Hz, and so on, about 155 cents upward.

(b) Treating the pharyngeal cavity as a stopped pipe resonating at 2.9 kHz:

$$L = \frac{v}{4f} = \frac{350\ \text{m/s}}{4(2900\ \text{Hz})} = 0.030\ \text{m} = 3.0\ \text{cm}. \tag{13.13}$$

Therefore the extra cavity is about 3 cm long, which is roughly the dimension of the widened pharynx just above a lowered larynx, and the reason the technique works.

Exercise 13.11

(Challenging) Show that a uniform tube closed at one end has resonances at odd multiples of $v/4L$, and explain why the vocal tract's formants are not *exactly* at these values.

Solution to Exercise 13.11

The closed end requires a displacement node and the open end a displacement antinode. Node and antinode are separated by a quarter wavelength, so the shortest fit is $L = \lambda/4$, and further fits add half-wavelengths:

$$L = (2m-1)\frac{\lambda}{4}, \qquad m = 1, 2, 3, \ldots \tag{13.14}$$

so $\lambda_m = 4L/(2m-1)$ and

$$f_m = \frac{v}{\lambda_m} = (2m-1)\frac{v}{4L}. \tag{13.15}$$

Three reasons the real tract departs from this:

1. **It is not uniform.** The tract's cross-section varies enormously along its length, and that variation is precisely what distinguishes one vowel from another. A uniform tube describes only the neutral vowel.

2. **The end correction applies** (Chapter 11). The lips are an open end of substantial radius, so the effective length exceeds the anatomical one.

3. **The nasal cavity can be coupled in** by lowering the soft palate, which adds a side branch with its own resonances and, more importantly, its own *anti*-resonances.

Therefore the stopped-pipe model gives the right ballpark and the right ordering, and everything interesting about vowels is a departure from it.

Exercise 13.12

(Straightforward) A singer moves from chest voice to falsetto. (a) Describe the change in fold vibration. (b) Predict the effect on the spectrum. (c) Explain why the transition is audible if untrained.

Solution to Exercise 13.12

(a) In chest voice the folds vibrate through their full depth and close completely on each cycle. In falsetto only the thin upper edges vibrate, and closure is often incomplete.

(b) Complete, abrupt closure produces a sharp discontinuity in the airflow, which by Chapter 5 means strong high harmonics. Incomplete, gentler closure produces a smoother waveform and a spectrum that rolls off much faster, plus added breath noise where air leaks through continuously.

So falsetto is spectrally poorer and breathier, and this is audible as a change of timbre quite apart from any change in pitch.

(c) At the transition, the mode of vibration changes discontinuously, so the spectrum changes discontinuously. The listener hears a sudden shift in tone quality, often with a momentary instability or crack as the folds reorganize.

Therefore training consists of learning to blend the two modes, mixing partial fold depth with adjusted tension, so that the spectral change is spread over several semitones instead of happening at one note.

Exercise 13.13

(Moderate) Explain why the voice is acoustically more like a violin than like a clarinet, despite being a wind-driven instrument.

Solution to Exercise 13.13

The distinction is which stage sets the pitch.

In a **clarinet** (Chapter 11), the bore's impedance peaks control the reed. The reed's own natural frequency is a couple of kilohertz and irrelevant; it oscillates at whatever the resonator dictates. Change the fingering and the pitch changes without touching the reed.

In a **violin** (Chapter 10), the string sets the pitch and the body merely responds. The body could be replaced with a different one and the same notes would sound, differently colored.

In the **voice**, the folds set the pitch, determined by their own mass and tension, and the tract merely shapes the result. Change the tract and the vowel changes; the pitch does not. The reason is impedance: the folds are massive and stiff enough that the tract's pressure oscillations cannot push them around.

Therefore the voice belongs with the violin: source sets pitch, resonator shapes timbre. Its wind-driven appearance is misleading, and the source–filter model is exactly the statement that the two stages are separable, which is what fails in a clarinet.

Exercise 13.14

(Moderate) A speech synthesizer produces vowels with correct formants and a perfectly steady pitch, and listeners describe it as robotic. Name three additions that would improve it, drawing on this chapter and Chapter 5.

Solution to Exercise 13.14

1. **Pitch variation.** Real speech has continuous intonation contours and small random fluctuations (jitter) of a fraction of a percent. A perfectly steady fundamental is the single strongest cue that a voice is synthetic, because no biological oscillator is that stable.

2. **Breath noise.** Real phonation leaks air continuously, adding a noise component with a continuous spectrum that no harmonic synthesis contains (Chapter 5). Its absence is what makes synthetic speech sound clean in an unnatural way.

3. **Moving formants.** Formants do not jump between target values; they glide, and the glides between vowels, the *transitions*, carry much of the information that identifies neighboring consonants. A synthesizer that hits formant targets and holds them produces separately correct vowels and unnatural speech.

A fourth worth adding is **amplitude variation**: real speech varies in level constantly with stress and syllable structure.

Therefore what is missing is again everything the steady-state model leaves out, which is the same conclusion Chapter 5 reached about synthesized instruments, and for the same reason.

Part VI

Sound in Space and in Circuits

Chapter 14

Room Acoustics and Concert-Hall Design

Learning Objectives

By the end of this chapter, you should be able to:

- Distinguish the direct sound, the early reflections, and the reverberant field, and sketch the echogram of a room.
- Explain why the inverse-square law fails indoors, and define the critical distance at which the reverberant field equals the direct sound.
- Define the absorption coefficient, calculate the total absorption of a room from its surfaces, and explain why absorption depends on frequency.
- Apply the Sabine equation to estimate reverberation time from room volume and absorption, and state the assumptions under which it is valid.
- State the reverberation times appropriate to speech, chamber music, orchestral music, and organ music, and explain why they differ.
- Explain the roles of specular reflection, diffusion, and absorption in a hall, and identify the acoustic defects each is used to address.
- Calculate the lowest room modes of a rectangular room, and explain why they matter in a small room and not in a large one.
- Identify the objective measures that correlate with the subjective judgment of a hall – early decay time, clarity, and lateral energy – and state what each measures.
- Compare the shoebox, fan, and vineyard hall plans, and explain the acoustic argument for and against each.

Introduction

Everything so far has treated the instrument as the source of the sound a listener hears. It is not. Between the instrument and the ear is a room, and in almost every case **most of what reaches the listener has bounced off something.**

In a concert hall, a listener in the middle of the stalls receives perhaps a tenth of their sound energy directly from the stage. The rest arrives later, from every direction, having reflected once, twice, or a hundred times. The room is not a transparent medium through which music passes; it is part of the instrument.

This has two consequences that run through the chapter.



Figure 14.1: The Copenhagen Concert Hall, its seating rising around the stage on every side rather than facing it from one end. §14.5 returns to why a "vineyard" plan like this one is one of three basic answers to the same design problem. Jacob Bøtter, CC BY 2.0.

The room is unavoidable. A performance in a different hall is a different performance, in a way a listener notices immediately. Composers have written for particular acoustics, Gabrieli for the reverberation of San Marco, chamber composers for small dry rooms, and a piece played in the wrong acoustic can be incoherent.

The room is designable, but only partly. The quantities that matter have been measurable for a century and the correlations with subjective judgment are good but not perfect. Halls built on the best available understanding have failed, and the failures have generally taught more than the successes.

The chapter builds three things in order: a picture of what arrives at a listener and when; the single most useful number, the reverberation time; and then the harder question of why some halls with the right reverberation time are nevertheless bad.

14.1 Sound in an Enclosure

Direct, Early, and Reverberant

A short sound on stage produces, at a listener's ear, a long and structured sequence of arrivals.

The three parts do different perceptual jobs.

The **direct sound** tells you where the source is. It arrives first, and the auditory system's precedence effect locates the source by it, disregarding the later arrivals for the purpose of direction.

The **early reflections** are *not* heard as echoes. Anything arriving within roughly 50–80 ms of the direct sound is fused with it, this is the **Haas effect**, and it makes the sound louder and fuller rather than doubled. Early reflections are therefore free loudness, and §14.4 shows that where they come *from* matters a great deal.

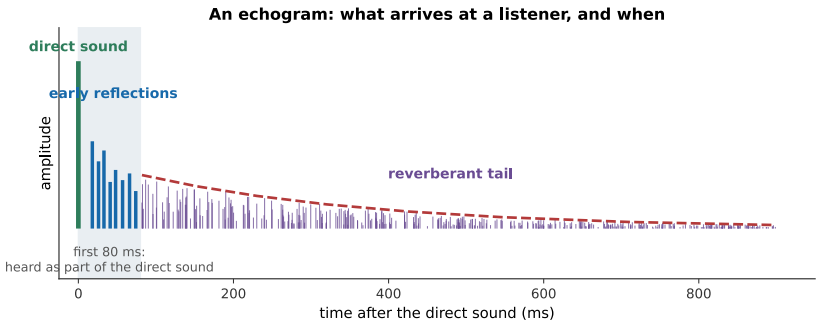


Figure 14.2: An echogram. The **direct sound** arrives first, by the shortest path. **Early reflections** follow within about 80 ms, having bounced once or twice. The **reverberant tail** is the accumulation of everything else, arriving so densely and from so many directions that individual reflections can no longer be distinguished.

The **reverberant tail** provides the sense of a space. It is what makes a hall sound large, and it is what a recording engineer adds artificially when a recording sounds dead.

The Echogram

The echogram is the room's **impulse response**: what a listener at one position receives from a source at one position, in response to a single impulse.

It contains everything linear about the room. Convolve any dry signal with it and you obtain what that signal would sound like at that seat, which is exactly how the audio examples in this chapter were made, and exactly how convolution reverb works in a recording studio (Chapter 15).

Critical Distance

Chapter 2 noted that the inverse-square law fails indoors. The place where it fails is the **critical distance**: the distance at which the direct sound and the reverberant field are equally strong.

$$r_c \approx 0.057\sqrt{\frac{V}{T_{60}}} \tag{14.1}$$

with the prefactor 0.057 in $\sqrt{\text{s/m}}$, V in cubic meters, and T_{60} in seconds.

Inside r_c, moving closer makes a real difference. Beyond it, the reverberant field dominates and moving further away changes the level very little. That is how the back row of a large hall stays audible at all, and also why it is so much less *clear* than the front.

Worked example: where the back row stops getting quieter

A hall of 20,000 m^3 *has* $T_{60} = 2.0$ *s. Find the critical distance, and comment.*

$$r_c = \left(0.057\sqrt{\frac{\text{s}}{\text{m}}}\right)\sqrt{\frac{20000\ \text{m}^3}{2.0\ \text{s}}} = \left(0.057\sqrt{\frac{\text{s}}{\text{m}}}\right)\sqrt{10000\ \text{m}^3/\text{s}} = 5.7\ \text{m}. \tag{14.2}$$

Only the first few rows are inside it. Everyone else is hearing a sound field dominated by reflections, at a level that barely changes from row ten to row forty.

> This is the central fact of concert-hall listening. It is why a hall can seat two thousand people without the back half straining to hear, and it is also why the back half hears something appreciably less clear than the conductor does.

14.2 Absorption and Reverberation Time

The Absorption Coefficient

Every reflection loses energy. The **absorption coefficient** α of a surface is the fraction of incident energy it does not reflect: 0 for a perfect mirror, 1 for an open window.

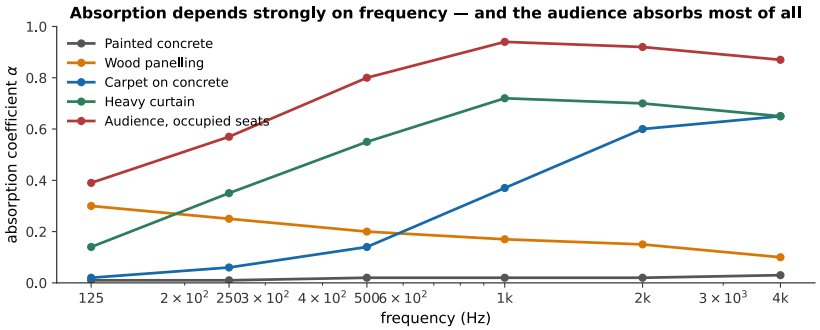


Figure 14.3: Absorption coefficients of common surfaces. Two things matter here. Absorption depends strongly on **frequency**, most soft materials absorb treble far better than bass, so untreated rooms tend to sound boomy. And **the audience is the most absorbent thing in the hall**.

That last point is a practical problem rather than a curiosity. A full hall and an empty hall are acoustically different rooms, and a rehearsal in an empty hall is a poor guide to the concert. Halls are designed with heavily upholstered seats chosen so that an empty seat absorbs approximately as much as an occupied one: concert-hall seating is therefore much plusher than cinema seating, and it is a specified acoustic component rather than a comfort decision.

The Sabine Equation

Wallace Sabine, working at Harvard in the 1890s on a lecture room that nobody could hear in, established the first quantitative result in architectural acoustics.

The **reverberation time** T_{60} is the time for the sound level to fall by 60 dB after the source stops: a factor of a million in energy. Sabine found

$$T_{60} = \frac{0.161\,V}{A}, \qquad A = \sum_i \alpha_i S_i, \tag{14.3}$$

with V the volume in cubic meters, S_i each surface's area, A the total absorption in square meters of equivalent open window, and the constant 0.161 in s/m (it is $55.3/v$ at room temperature; see Exercise 14.11).

The equation says something simple: **bigger rooms ring longer, more absorbent rooms ring less**, and the trade is linear in both.

Warning

The Sabine equation assumes a **diffuse** field, sound arriving equally from all directions and absorption spread evenly over the surfaces. It fails when a room is very absorbent (an anechoic chamber's predicted T_{60} is nonsense), when absorption is concentrated on one surface, and when the room is long and thin. The Eyring and Millington refinements address the first case; nothing addresses a badly shaped room except not building it.

How Long Should a Hall Ring?

There is no single right answer, because different music wants different things.

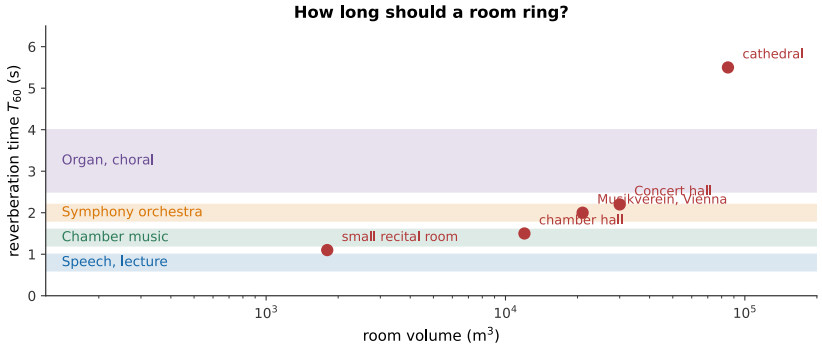


Figure 14.4: Reverberation times for different purposes. Speech needs a short time, because syllables arriving on top of one another destroy intelligibility. Orchestral music wants around 2 s. Organ and choral music, written for buildings with 4 s or more, want the reverberation to be part of the texture.

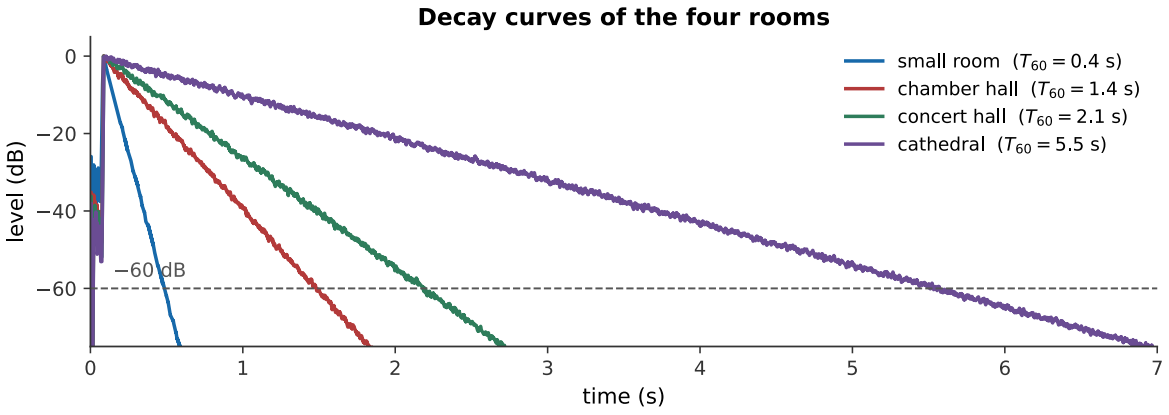


Figure 14.5: One phrase, four rooms. Listen to the *last* note in each: in the cathedral it is still audible long after the phrase has ended, and the earlier notes are still sounding underneath it. That blurring is why music written for such buildings moves slowly and changes harmony rarely. *What you hear:*The same four-note phrase four times. The first is completely dry and sounds artificial and close. The second is natural but small. The third is full and spacious. The fourth blurs the notes into each other so that the last chord contains all of them. Listen: Dry (anechoic), Small room, Concert hall, Cathedral

The conflict between speech and music is real and unresolvable. A hall optimized for an

orchestra is poor for a lecture and vice versa, and a multi-purpose hall is a compromise that satisfies nobody completely. Section 14.5 therefore takes up variable acoustics.



Figure 14.6: The same saxophone is heard in two acoustic extremes: first surrounded by a long reverberant tail, then almost entirely stripped of reflected sound. The contrast isolates how strongly a room becomes part of an instrument's sound. Watch on YouTube: Saxophone in a Reverberation Room and an Anechoic Chamber

14.3 Reflection, Diffusion, and Modes

Specular Reflection and Its Defects

A flat, hard surface reflects sound **specularly**: angle in equals angle out, like a mirror. Useful for directing sound where you want it, and the source of three classic defects.

An **echo** is a single reflection arriving late enough, more than about 80 ms after the direct sound, so more than about 27 m of extra path, and strongly enough to be heard separately. A flat rear wall in a long hall is the usual culprit.

Flutter echo is a rapid series of reflections bouncing between two parallel hard surfaces, heard as a metallic ringing after a handclap. Splaying one wall by a few degrees eliminates it entirely.

Focusing occurs when a concave surface, a domed ceiling, a curved rear wall, concentrates reflections at a point. The focus is much too loud and the surrounding seats are starved. Con-

cave surfaces facing an audience are avoided for this reason, and where architecture demands one it is usually broken up or covered.

Diffusion

The remedy for all three is **diffusion**: surfaces that scatter sound in many directions rather than reflecting it in one.

Diffusion is why historic halls sound as good as they do. The Musikverein in Vienna and the Concertgebouw in Amsterdam are covered in statuary, coffered ceilings, niches, balustrades, and moldings, all on the scale of a few tens of centimeters, comparable to the wavelengths of musical sound, and therefore excellent scatterers. The decoration was not acoustically motivated, and it turned out to be acoustically essential.

Twentieth-century halls with large smooth surfaces often disappointed for exactly this reason, and modern designs restore diffusion deliberately, sometimes using mathematically designed diffusers whose depths follow a number-theoretic sequence.

Room Modes and Why Small Rooms Are Hard

In a large hall the sound field is dense enough to treat statistically. In a small room it is not.

A rectangular room's modes are at

$$f_{lmn} = \frac{v}{2}\sqrt{\left(\frac{l}{L_x}\right)^2 + \left(\frac{m}{L_y}\right)^2 + \left(\frac{n}{L_z}\right)^2}, \tag{14.4}$$

and at low frequencies they are **sparse**.

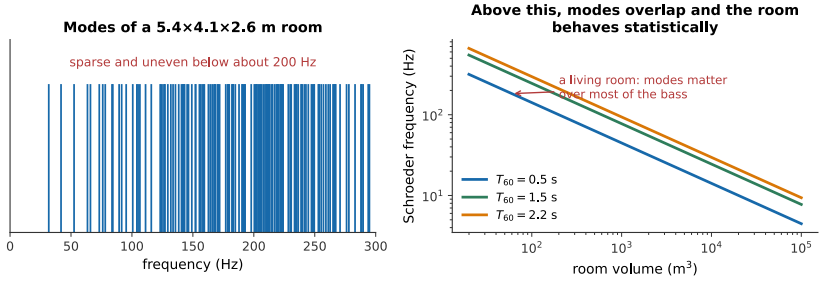


Figure 14.7: Left: the modes of a small room, individually resolvable below about 200 Hz. Right: the **Schroeder frequency**, above which modes overlap enough that the room behaves statistically. For a living room it is a few hundred hertz; for a concert hall it is around 20 Hz. Halls can therefore be analyzed statistically; bedrooms cannot.

Below the Schroeder frequency the room's response is a sparse set of peaks and nulls that does not average out. Bass notes at a modal frequency boom; bass notes between modes disappear. Moving your head changes the bass. This is the fundamental difficulty of small-room acoustics, and a studio control room costs more to treat than a concert hall costs per cubic meter for just that reason.

14.4 What Makes a Hall Sound Good

Beyond Reverberation Time

By the mid-twentieth century T_{60} was well understood, and halls were being built with correct reverberation times that musicians disliked. Something was missing.

What was missing was that T_{60} is a single number describing an average over the whole room and the whole decay, and listeners respond to much more: to *when* the energy arrives, from *which direction*, and at *which frequencies*.

Clarity, Warmth, and Intimacy

Several measures now supplement T_{60}, and each corresponds to something a musician can name.

Early decay time (EDT) measures the slope of the first 10 dB of decay rather than extrapolating from 60. It corresponds far better to perceived reverberance, because the first 10 dB is what a listener actually hears before the next note arrives.

Clarity C_{80} is the ratio, in decibels, of energy arriving in the first 80 ms to everything after:

$$C_{80} = 10 \log_{10} \frac{\int_0^{80\,\mathrm{ms}} p^2 \,\mathrm{d}t}{\int_{80\,\mathrm{ms}}^{\infty} p^2 \,\mathrm{d}t}. \tag{14.5}$$

High C_{80} means detail and articulation; low means blend and wash. Around 0 dB suits orchestral music; higher suits opera, where words matter.

Warmth (bass ratio) compares reverberation time at low frequencies to that at mid frequencies. A hall whose bass rings longer sounds warm; one whose bass dies first sounds thin. Since Chapter 14's absorption figure showed that soft materials absorb bass *worst*, warmth is usually easy to obtain and thinness usually means too much paneling.

Intimacy is the delay between the direct sound and the first reflection. A short delay makes a large hall feel small, and it is the reason side walls close to the audience matter.

Lateral Energy and Spatial Impression

The most important discovery of twentieth-century hall acoustics is that **the direction the early reflections come from matters more than their strength**.

Reflections arriving from the **sides** make the sound seem to come from a broad source and make the listener feel surrounded. Reflections from the **ceiling** do not, because they reach both ears nearly identically.

The reason is binaural: a lateral reflection reaches the two ears at different times and levels, and that difference is what produces the impression of spaciousness. A ceiling reflection is symmetric and the auditory system simply folds it into the direct sound.

The measure is the **lateral fraction**, the proportion of early energy arriving from the sides, and it correlates with subjective preference better than almost anything else.

This single result explains the enduring success of the **shoebox** hall. A narrow rectangular room puts every listener close to a side wall; a wide fan-shaped one does not.

14.5 Halls in Practice

Shoebox, Fan, and Vineyard

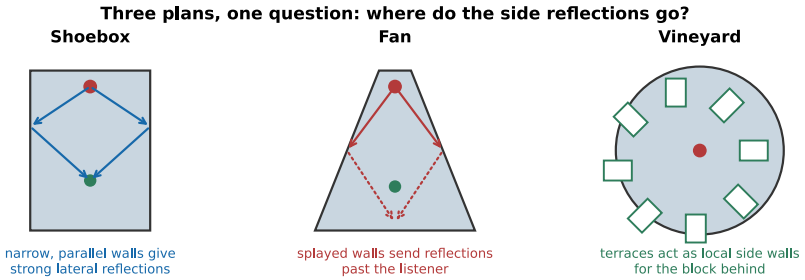


Figure 14.8: Three plans. The question in each case is where the early side reflections go.

The **shoebox**, Vienna's Musikverein (1870), Amsterdam's Concertgebouw (1888), Boston Symphony Hall (1900), is narrow, high, rectangular, and heavily decorated. It is consistently at the top of every ranking of concert halls, and the reasons are now understood: narrow width gives strong lateral reflections, high ceilings give volume without width, and the decoration diffuses.

Its disadvantage is capacity. A shoebox cannot be widened without losing what makes it work, so seats must go into balconies, and the hall becomes tall and expensive.

The **fan** shape spread in the twentieth century because it seats more people closer to the stage. Acoustically it is usually worse: splayed side walls reflect sound toward the *back* of the hall rather than across it, so the lateral fraction is low and the sound is clear but flat.

The **vineyard**, introduced at the Berlin Philharmonie (1963), surrounds the orchestra with terraced blocks of seating. The retaining walls of each terrace act as local side walls for the block behind, providing lateral reflections that a single large volume could not. It also allows large capacity with everyone close to the stage.

Famous Successes and Famous Failures

The literature's standard cautionary tale is the original Philharmonic Hall at New York's Lincoln Center (1962), which was widely judged a failure on opening, thin bass and poor blend, and was rebuilt more than once, eventually being gutted entirely.

The useful lesson is not that its designers were careless; they were not. It is that the quantities known to matter in 1962 were fewer than the quantities that do matter, and that a hall is a single very expensive experiment that cannot be run twice. Much of what is now known about lateral energy was learned from halls that had already been built.

Variable Acoustics and Electronic Enhancement

Since no fixed acoustic suits everything, modern halls increasingly vary theirs.

Mechanically: retractable curtains and banners over wall and ceiling absorbers; adjustable reverberation chambers, which are volumes coupled to the main hall through openings that can be closed; movable ceiling reflectors above the stage.

Electronically: a system of microphones, processors, and many distributed loudspeakers that adds reverberant energy. Modern systems are good enough that listeners cannot reliably

detect them, and they can turn a dry drama theater into a plausible concert hall at the press of a button.

Electronic enhancement remains controversial, and the argument is worth stating honestly. The objection is not that it does not work, it does, but that a hall's acoustic has been part of the compositional and performing tradition, and that a variable acoustic makes it a production decision rather than a fact about the building. That is an aesthetic disagreement rather than a physical one, and this book takes no side in it.



Figure 14.9: Students from the University of Edinburgh measure the Glasgow Royal Concert Hall. Their tests connect reverberation time and reflections to the practical work of characterizing a real performance space. Watch on YouTube: Glasgow Royal Concert Hall Acoustics

14.6 Summary

- **Most of what a listener hears has reflected.** The room is part of the instrument, and beyond the **critical distance** $r_c \approx 0.057\sqrt{V/T_{60}}$ the reverberant field dominates and the inverse-square law fails.
- **Three parts of the arrival**: the **direct sound** locates the source; **early reflections** within about 80 ms fuse with it and add loudness and fullness (the Haas effect); the **reverberant tail** gives the sense of space.
- **The echogram is the room's impulse response**, and convolving a dry signal with it reproduces what that seat would hear, which is how convolution reverb works.

- **The absorption coefficient** is frequency-dependent: soft materials absorb treble far better than bass. **The audience is the most absorbent element**, so seats are designed to absorb like occupied ones.
- **The Sabine equation** $T_{60} = 0.161V/A$ (with 0.161 in s/m) gives reverberation time from volume and total absorption. It assumes a diffuse field and fails for very absorbent, awkwardly shaped, or unevenly treated rooms.
- **Target reverberation times** run from 0.6–1.0 s for speech through 1.8–2.2 s for orchestral music to 4 s and beyond for organ and choral music.
- **Specular reflection produces echo, flutter, and focusing**; diffusion cures all three, and is why heavily decorated historic halls sound good.
- **Below the Schroeder frequency, room modes are individually audible.** This is the fundamental difficulty of small rooms and is why bass is position-dependent in a living room and not in a hall.
- T_{60} **is not enough.** Early decay time, clarity C_{80}, warmth, and intimacy each correspond to something musicians can name.
- **Lateral energy is the key discovery**: early reflections from the *sides* produce spaciousness because they reach the two ears differently, while ceiling reflections do not. Narrow shoebox halls outperform wide fan-shaped ones for that reason.

14.7 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple choice.** A diffuse room has volume 1000 m^3 and total absorption 100 m^2 sabins. Does the Sabine equation predict about 0.16, 1.6, 6.2, or 16 s?

2. **True or false.** Beyond the critical distance, level continues to fall by 6 dB for every doubling of distance because direct sound dominates.

3. **Drag the words.** Match direct sound, early reflections, the reverberant tail, and lateral energy to their perceptual roles.

4. **Fill in the blanks.** Beyond the _____ distance, the _____ field dominates over direct sound.

5. **Mark the words.** Identify three specular-reflection defects that diffusion helps cure: "Diffusion helps prevent distinct echo, repeated flutter, and curved-wall focusing by scattering reflections."

14.8 Conceptual Questions

1. Explain why a listener in the back row of a large hall hears a sound only a little quieter than one in row ten, but appreciably less clear.

2. Explain why a reflection arriving 30 ms after the direct sound is not heard as an echo, while one arriving 150 ms later is.

3. A hall sounds very different when empty and when full. Explain why, and say what hall designers do about it.

4. Explain why a room treated only with thin carpet and curtains tends to sound boomy.
5. A handclap in a corridor produces a metallic ringing. Name the defect, explain its cause, and give a cure.
6. Explain why room modes are a serious problem in a living room and not in a concert hall.
7. Explain why an early reflection from a side wall contributes more to a sense of spaciousness than an equally strong one from the ceiling.
8. Explain why fan-shaped halls seat more people and generally sound worse than shoebox halls.

14.9 Problems

Exercise 14.1
(Straightforward) A hall has a volume of 18,000 m^3 and a total absorption of 1450 m^2. (a) Find its reverberation time. (b) Is this suitable for orchestral music?

Solution to Exercise 14.1
(a) Using the Sabine equation:

$$T_{60} = \frac{0.161V}{A} = \frac{(0.161\ \mathrm{s/m})(18000\ \mathrm{m}^3)}{1450\ \mathrm{m}^2} = 2.00\ \mathrm{s}. \tag{14.6}$$

(b) Yes. Two seconds sits squarely in the 1.8–2.2 s band that orchestral music wants, as Figure 14.4 shows.

Exercise 14.2
(Moderate) A rectangular hall is 40 m by 22 m by 16 m. Its surfaces have an average absorption coefficient of 0.15. (a) Find the volume. (b) Find the total surface area. (c) Find the reverberation time.

Solution to Exercise 14.2
(a) $V = 40 \times 22 \times 16 = 1.41 \times 10^4\ \mathrm{m}^3$.
(b) Two of each pair of faces:

$$S = 2(40\times 22) + 2(40\times 16) + 2(22\times 16) = 1760 + 1280 + 704 = 3744\ \mathrm{m}^2. \tag{14.7}$$

(c) $A = \alpha S = 0.15(3744) = 562\ \mathrm{m}^2$, so

$$T_{60} = \frac{(0.161\ \mathrm{s/m})(14080\ \mathrm{m}^3)}{562\ \mathrm{m}^2} = 4.03\ \mathrm{s}. \tag{14.8}$$

Therefore about 4 s, far too long for orchestral music. The hall would need its absorption roughly doubled, which is what an audience and upholstered seating provide.

Exercise 14.3
(Moderate) The hall of Exercise 14.2 is fitted with 1200 seats, each with an absorption of 0.45 m^2 when empty and 0.85 m^2 when occupied. (a) Find T_{60} empty. (b) Find T_{60} full. (c) Comment on rehearsing there.

Solution to Exercise 14.3
(a) Empty: $A = 562 + 1200(0.45) = 562 + 540 = 1102\ \mathrm{m}^2$:

$$T_{60} = \frac{(0.161\ \mathrm{s/m})(14080\ \mathrm{m}^3)}{1102\ \mathrm{m}^2} = 2.06\ \mathrm{s}. \tag{14.9}$$

(b) Full: $A = 562 + 1200(0.85) = 562 + 1020 = 1582\ \mathrm{m}^2$:

$$T_{60} = \frac{(0.161\ \mathrm{s/m})(14080\ \mathrm{m}^3)}{1582\ \mathrm{m}^2} = 1.43\ \mathrm{s}. \tag{14.10}$$

(c) The hall is 0.6 s longer empty: a difference every musician would notice. An orchestra rehearsing in the empty hall hears more reverberation than the audience will, so it will tend to play with less sustain and less blend than the concert requires, and balances set in rehearsal will be wrong.

This is exactly why hall designers specify seats whose empty absorption is close to their occupied absorption. Here the ratio is 0.45 to 0.85, which is poor; a well-designed seat would be nearer 0.75 to 0.85.

Exercise 14.4
(Moderate) A hall has $V = 25{,}000\ \mathrm{m}^3$ and $T_{60} = 2.2$ s. (a) Find the critical distance. (b) A listener moves from 8 m to 32 m from the stage. How much does the level drop, and how much would it drop outdoors?

Solution to Exercise 14.4
(a)

$$r_c = \left(0.057\sqrt{\frac{\mathrm{s}}{\mathrm{m}}}\right)\sqrt{\frac{25000\ \mathrm{m}^3}{2.2\ \mathrm{s}}} = \left(0.057\sqrt{\frac{\mathrm{s}}{\mathrm{m}}}\right)\sqrt{11364\ \mathrm{m}^3/\mathrm{s}} = 6.1\ \mathrm{m}. \tag{14.11}$$

(b) Both positions are beyond the critical distance, so both are in the reverberant field, where the level is essentially uniform. The drop is small: a couple of decibels at most.

Outdoors, the inverse-square law gives

$$20\log_{10}\left(\frac{32}{8}\right) = 12\ \mathrm{dB}. \tag{14.12}$$

Therefore the room saves the distant listener about 10 dB, which is most of the reason a two-thousand-seat hall works without amplification.

Exercise 14.5
(Moderate) A reflection travels a path 19 m longer than the direct sound. (a) What is the delay? (b) Is it heard as an echo or fused with the direct sound? (c) What path difference would make it an echo?

Solution to Exercise 14.5

(a) $t = 19/343 = 0.055$ s $= 55$ ms.

(b) Fused. Anything within about 80 ms is integrated with the direct sound by the Haas effect, and contributes loudness rather than being heard separately.

(c) The threshold is about 80 ms, so

$$d = vt = (343 \text{ m/s})(0.080 \text{ s}) = 27 \text{ m} \tag{14.13}$$

of extra path. In a hall 40 m long, a listener near the front and a hard rear wall can easily produce this, so rear walls are treated.

Exercise 14.6

(Moderate) A room is 6.0 m by 4.5 m by 2.8 m. (a) Find the three lowest axial modes. (b) Find the Schroeder frequency if $T_{60} = 0.5$ s. (c) Comment on bass reproduction.

Solution to Exercise 14.6

(a) Axial modes are at $f = v/2L$ along each dimension:

$$\frac{343 \text{ m/s}}{2(6.0 \text{ m})} = 28.6 \text{ Hz}, \quad \frac{343 \text{ m/s}}{2(4.5 \text{ m})} = 38.1 \text{ Hz}, \quad \frac{343 \text{ m/s}}{2(2.8 \text{ m})} = 61.3 \text{ Hz}. \tag{14.14}$$

(b) $V = 6.0 \times 4.5 \times 2.8 = 75.6$ m^3:

$$f_S = \left(2000 \text{ Hz}\sqrt{\frac{\text{m}^3}{\text{s}}}\right)\sqrt{\frac{0.5 \text{ s}}{75.6 \text{ m}^3}} = \left(2000 \text{ Hz}\sqrt{\frac{\text{m}^3}{\text{s}}}\right)\left(0.0813\sqrt{\frac{\text{s}}{\text{m}^3}}\right) = 163 \text{ Hz}. \tag{14.15}$$

(c) Below 163 Hz the modes are individually resolvable, and the three lowest are widely separated at 29, 38, and 61 Hz. A bass note landing on a mode will boom; one landing between modes will nearly disappear; and both effects depend on where the listener sits.

Therefore bass in this room is position-dependent and uneven over the whole of the lowest two octaves of music, which is the normal state of domestic listening, and the reason studio control rooms need extensive low-frequency treatment.

Exercise 14.7

(Moderate) A lecture room of 600 m^3 has $T_{60} = 1.8$ s and is unintelligible. (a) How much absorption does it currently have? (b) How much is needed for $T_{60} = 0.8$ s? (c) If acoustic panels have $\alpha = 0.85$, what area is needed?

Solution to Exercise 14.7

(a) From $A = 0.161V/T_{60}$:

$$A = \frac{(0.161 \text{ s/m})(600 \text{ m}^3)}{1.8 \text{ s}} = 53.7 \text{ m}^2. \tag{14.16}$$

(b) For 0.8 s:

$$A = \frac{(0.161 \text{ s/m})(600 \text{ m}^3)}{0.8 \text{ s}} = 120.8 \text{ m}^2. \tag{14.17}$$

(c) The extra absorption needed is $120.8 - 53.7 = 67.1$ m^2. Panels of $\alpha = 0.85$ provide 0.85 m^2 of absorption per square meter, but they also cover an existing surface. Taking the existing wall as roughly $\alpha = 0.05$, each square meter of panel adds a net 0.80 m^2:

$$\text{area} = \frac{67.1}{0.80} = 84 \text{ m}^2. \tag{14.18}$$

Therefore about 84 m^2 of paneling: a substantial fraction of the room's surface, and the reason retrofitting a bad lecture room is expensive. This is essentially the problem Sabine was hired to solve.

Exercise 14.8
(Moderate) A hall has $T_{60} = 2.4$ s at 125 Hz and 1.9 s at 1 kHz. (a) Find the bass ratio. (b) Is the hall warm or thin? (c) What would raise the bass ratio?

Solution to Exercise 14.8
(a) The bass ratio is the low-frequency reverberation time divided by the mid:

$$\frac{2.4}{1.9} = 1.26. \tag{14.19}$$

(b) **Warm.** A bass ratio above about 1.1–1.2 is the usual criterion for warmth, and this hall exceeds it.

(c) Anything that absorbs more mid and high frequency or less bass. In practice the lever is usually the other way round: to *lower* a bass ratio you add bass traps, since Chapter 14's absorption figure shows that ordinary soft furnishing absorbs bass poorly. To raise it further one would reduce thin paneling, which acts as a low-frequency absorber by flexing, and use massive rigid surfaces instead, which is what the stone and plaster of historic halls provide.

Exercise 14.9
(Challenging) Explain quantitatively why a listener 30 m from an orchestra hears mostly reflected sound, in a hall with $V = 20{,}000$ m^3 and $T_{60} = 2.0$ s.

Solution to Exercise 14.9
The critical distance is

$$r_c = \left(0.057\sqrt{\frac{\text{s}}{\text{m}}}\right)\sqrt{\frac{20000 \text{ m}^3}{2.0 \text{ s}}} = 5.7 \text{ m}. \tag{14.20}$$

By definition, at r_c the direct and reverberant energies are equal. The reverberant energy is uniform throughout the room, while the direct energy falls as $1/r^2$. So at 30 m the direct sound has fallen relative to its value at r_c by

$$\left(\frac{5.7}{30}\right)^2 = 0.036, \tag{14.21}$$

while the reverberant energy is unchanged.

Therefore the direct sound is only about 3.6% of the total: the listener is receiving roughly 96% reflected energy. In decibels the direct sound is $10\log_{10}(0.036) = -14$ dB relative to the reverberant field.

This is the quantitative form of the chapter's opening claim, and it is why the room matters as much as the orchestra.

Exercise 14.10
(Moderate) A hall's impulse response has 2.4 units of energy in the first 80 ms and 3.1 units afterwards. (a) Find C_{80}. (b) Is this hall better suited to symphonic music or to opera? (c) What change would raise C_{80}?

Solution to Exercise 14.10
(a)

$$C_{80} = 10\log_{10}\left(\frac{2.4}{3.1}\right) = 10\log_{10}(0.774) = -1.1 \text{ dB}. \tag{14.22}$$

(b) **Symphonic music.** A slightly negative C_{80} means the late energy slightly exceeds the early, which gives blend and fullness at some cost in articulation, good for orchestral texture. Opera wants words to be intelligible and is usually happier with C_{80} around +1 to +4 dB.

(c) Raising C_{80} means shifting energy earlier. In practice: reflectors placed to return sound to the audience quickly, a lower ceiling over the stage, and more absorption on distant surfaces to shorten the tail. Reducing the overall volume does both at once.

Exercise 14.11
(Challenging) Derive the Sabine equation's form, given that sound energy in a room decays exponentially with the mean time between reflections and the fraction absorbed at each.

Solution to Exercise 14.11
In a diffuse field, the mean free path between reflections is $4V/S$, where S is the total surface area. The mean time between reflections is therefore

$$\tau = \frac{4V}{Sv}. \tag{14.23}$$

At each reflection a fraction $\bar{\alpha}$ of the energy is absorbed, so after n reflections the energy is $E_0(1-\bar{\alpha})^n$. Writing $n = t/\tau$:

$$E(t) = E_0\,(1-\bar{\alpha})^{\,t/\tau} = E_0\exp\!\left(\frac{t}{\tau}\ln(1-\bar{\alpha})\right). \tag{14.24}$$

For small $\bar{\alpha}$, $\ln(1-\bar{\alpha}) \approx -\bar{\alpha}$, so

$$E(t) = E_0\exp\!\left(-\frac{\bar{\alpha}Sv\,t}{4V}\right). \tag{14.25}$$

Set $E/E_0 = 10^{-6}$, which is -60 dB:

$$\frac{\bar{\alpha}Sv\,T_{60}}{4V} = \ln(10^6) = 13.82, \tag{14.26}$$

$$T_{60} = \frac{55.3\,V}{v\,\bar{\alpha}S} = \frac{55.3\,V}{(343\text{ m/s})\,A} = \frac{(0.161\text{ s/m})\,V}{A}. \tag{14.27}$$

Therefore the constant 0.161 s/m is 55.3/343, it contains the speed of sound, and would differ in another medium or at another temperature. Note also where the small-$\bar{\alpha}$ approximation entered: it is exactly why the equation fails for very absorbent rooms, and the Eyring correction is what you get by keeping $\ln(1-\bar{\alpha})$ intact.

Exercise 14.12
(Challenging) Two halls have identical T_{60} of 2.0 s. One is a 22 m wide shoebox and the other a 42 m wide fan. (a) Estimate the delay of the first side-wall reflection for a listener on the center line, 20 m from the stage, in each. (b) Which sounds more spacious, and why?

Solution to Exercise 14.12
(a) Take the source on the center line at the stage and the listener 20 m back. For a wall at distance $w/2$ from the center line, the reflected path is approximately

$$2\sqrt{\left(\frac{w}{2}\right)^2 + \left(\frac{20}{2}\right)^2}. \tag{14.28}$$

Shoebox, $w = 22$ m: path $= 2\sqrt{11^2 + 10^2} = 2(14.87) = 29.7$ m, against 20 m direct: an extra 9.7 m, so a delay of 28 ms.

Fan, $w = 42$ m: path $= 2\sqrt{21^2 + 10^2} = 2(23.26) = 46.5$ m: an extra 26.5 m, so a delay of 77 ms.

(b) The **shoebox**. Its side reflections arrive at 28 ms, comfortably inside the fusion window, strongly, and from the sides, exactly the lateral energy that produces spatial impression. The fan's arrive at 77 ms, at the very edge of the window, much weaker for having traveled further, and from a shallower angle that is closer to frontal than lateral.

Therefore the same reverberation time can produce very different halls, which is the whole point of §14.4 and the reason T_{60} alone was never enough.

Exercise 14.13
(Moderate) A recording engineer wants to place an instrument in a virtual hall using convolution reverb. (a) What must they measure in the real hall? (b) What operation do they perform? (c) What does this assume about the room?

Solution to Exercise 14.13
(a) The **impulse response** of the hall, from the source position to the listening position, in practice measured with a swept sine or a starter pistol and a microphone, then processed to recover the response.

(b) **Convolution** of the dry recording with the impulse response. Every sample of the dry signal is replaced by a scaled copy of the whole impulse response, and the copies are summed.

(c) That the room is **linear** and **time-invariant**. Linear means superposition holds (Chapter 3), so the response to a sum of sounds is the sum of the responses, true for any room at musical levels. Time-invariant means the room does not change during the recording, which fails if the audience arrives, the temperature drifts, or a door opens.

It also assumes the source and listener stay where they were when the response was measured, so a moving source in a convolution reverb never sounds quite right.

Exercise 14.14
(Moderate) A multi-purpose hall must serve both speech ($T_{60} = 0.9$ s) and orchestral music ($T_{60} = 2.0$ s). Its volume is 12,000 m^3. (a) What absorption does each require? (b) How much absorption must be removable? (c) Suggest two mechanisms.

Solution to Exercise 14.14

(a) For speech:

$$A = \frac{(0.161\ \text{s/m})(12000\ \text{m}^3)}{0.9\ \text{s}} = 2147\ \text{m}^2. \tag{14.29}$$

For music:

$$A = \frac{(0.161\ \text{s/m})(12000\ \text{m}^3)}{2.0\ \text{s}} = 966\ \text{m}^2. \tag{14.30}$$

(b) The difference is $2147 - 966 = 1181\ \text{m}^2$ of absorption that must be introduced for speech and removed for music, more than the entire absorption required in the music configuration.

(c) Two mechanisms:

- **Retractable absorbers**: heavy velour banners on motorized rollers, deployed over wall and ceiling surfaces. Roughly 1400 m^2 of banner at $\alpha \approx 0.85$ would do it, which is a great deal of fabric and storage volume.
- **Coupled reverberation chambers**: volumes adjoining the hall, connected through motorized doors. Opening them increases the effective volume and the reverberation; closing them shortens it. This is the more elegant solution, since it changes V rather than A, and it is used in several modern halls.

Therefore the required change is large, and a hall that does both well is a genuinely difficult and expensive building. Most venues therefore choose one purpose and accept being mediocre at the other.

Chapter 15

Electronic and Recorded Sound

Learning Objectives

By the end of this chapter, you should be able to:

- Describe how dynamic and condenser microphones convert sound pressure into a voltage, and compare their characteristics.
- Read a microphone polar pattern, and choose an appropriate pattern for a stated recording situation.
- Explain how a loudspeaker converts a signal back into sound, and why a single driver cannot cover the audible range.
- Explain how a magnetic pickup converts string motion into a voltage, and why pickup placement changes an electric guitar's timbre.
- Interpret a frequency-response curve, and distinguish linear distortion from nonlinear distortion.
- State the sampling theorem, calculate the Nyquist frequency for a given sampling rate, and explain why 44.1 kHz was chosen for the compact disc.
- Explain aliasing, predict the frequency of an aliased component, and describe the role of the anti-aliasing filter.
- Relate bit depth to dynamic range through the six-decibel-per-bit rule, and explain what dither does and why it helps.
- Explain how perceptual coding uses masking to discard information, and state what is and is not recoverable afterward.
- Compare additive, subtractive, FM, and sampling synthesis, and identify which chapter of the book each one is an application of.
- Explain how two loudspeakers produce a stereo image, and describe the cues a recording uses to place a source in space.

Introduction

Almost all the music anybody hears has been through a microphone, a wire, and a loudspeaker. Live acoustic performance is now the exception rather than the rule, and this chapter is about the chain that has replaced it.

It is a satisfying place to end, because very little in it is new. The chain turns out to be an application of everything already established:

- A **microphone** is a resonator with a diaphragm (Chapter 4) whose directivity is governed by size against wavelength (Chapter 2).



Figure 15.1: Recording begins with a membrane converting pressure into motion and then voltage. Shixart1985, CC BY 2.0.

- **Digital audio** rests on Fourier's theorem (Chapter 5): a signal with no energy above some frequency can be captured by samples taken fast enough.
- **Perceptual coding** works because masking (Chapter 7) makes much of a signal inaudible, so it need not be stored.
- **Synthesis** is the spectrum of Chapter 5 and the source–filter model of Chapter 13, built deliberately rather than found in an instrument.
- **Stereo** exploits the localization cues of Chapter 6, and artificial reverberation is the impulse response of Chapter 14.

There is one genuinely new idea, and it is a deep one: a continuous signal can be represented **exactly** by a finite list of numbers. The rest is engineering.



Figure 15.2: Technology Connections explains the 45/45 groove geometry that stores two audio channels in one mechanical trace. The construction is a compact example of turning two electrical signals into motion and back again. Watch on YouTube: How Do Vinyl Records Hold Stereo Sound?

15.1 Microphones and Loudspeakers

Turning Pressure Into Voltage

A microphone is a pressure detector with a diaphragm, and there are two common ways to read its motion.

A **dynamic** microphone attaches a coil of wire to the diaphragm and suspends it in a



Figure 15.3: Ben Krasnow images a record groove directly with an electron microscope. The magnified walls make the physical waveform traced by a playback stylus visible rather than abstract. Watch on YouTube: Vinyl LP in an Electron Microscope

magnetic field. Motion induces a voltage. It is rugged, needs no power, and handles very loud sources, but the coil's mass limits its high-frequency response and its transient accuracy.

A **condenser** microphone makes the diaphragm one plate of a capacitor. Motion changes the capacitance and hence the voltage across it. The diaphragm can be made extremely light, so the response extends further and transients are tracked more faithfully, at the cost of needing a power supply and being more fragile.

The mechanical requirement is the same in both cases and is worth stating: the diaphragm's resonance should be placed *outside* the band of interest, and it should be heavily damped, so that the microphone responds evenly rather than coloring everything with its own resonance. This is Chapter 4's low-Q design philosophy, for the same reason a guitar body needs it.

Polar Patterns

A microphone's directivity is set by whether its diaphragm is exposed to pressure on one side or both.

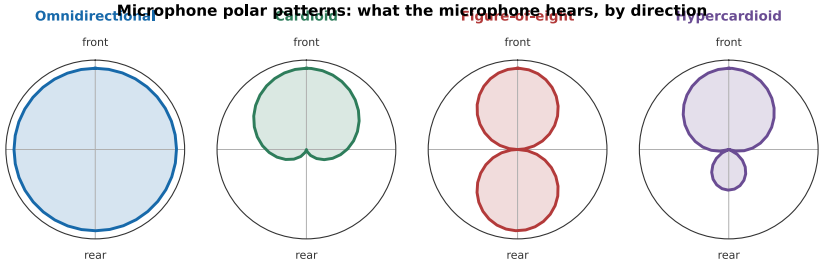


Figure 15.4: The four standard patterns. **Omnidirectional** responds to pressure alone and hears everything equally. **Figure-of-eight** responds to the pressure *difference* across the diaphragm, so it is deaf at the sides. **Cardioid** is the sum of the two, giving a rear null. **Hypercardioid** is a different mixture, trading a small rear lobe for a narrower front.

The choice is a practical one. A cardioid rejects the room and the audience, which suits a close-miked instrument in a poor room. An omnidirectional captures the room, which suits a good hall. A figure-of-eight rejects sound from the sides, and is used to isolate two instruments facing each other.

> **Note**
> Any directional microphone, that is, any that responds to pressure *difference*, exhibits **proximity effect**: the bass response rises sharply as the source comes close, by 10 dB or more within a few centimeters. Broadcasters exploit it deliberately for an authoritative voice, and singers who work close to the microphone must either compensate or lean back.

Loudspeakers, Drivers, and Crossovers

A loudspeaker runs the microphone backwards: current through a coil in a magnetic field produces force, which moves a cone, which pushes air.

It faces a problem Chapter 2 identified. Directivity depends on the source's size compared with the wavelength, and a single driver has one size while music spans a factor of a thousand in wavelength. A driver large enough to move air at 40 Hz beams badly at 4 kHz; one small enough to radiate 10 kHz evenly cannot shift enough air to produce bass at all.

The solution is multiple drivers, a **woofer**, sometimes a midrange, and a **tweeter**, with a **crossover** network dividing the signal between them. The crossover is itself a compromise, since the drivers must blend through the handover region without canceling, and crossover design is much of what distinguishes loudspeakers.

The low-frequency problem is attacked separately. A cone radiates from both faces, in opposite phase, so at long wavelengths the front and rear outputs meet around the edge and cancel. An enclosure prevents that. A **sealed** box is simple and well-behaved; a **bass-reflex** box adds a tuned port, a Helmholtz resonator (Chapter 4), that reinforces the bottom octave at the cost of a steeper roll-off below it.



Figure 15.5: Branch Education follows an electrical signal through the magnet and voice coil to the moving cone and the pressure wave radiated into the room. The cutaway view shows how the components described above fit together. Watch on YouTube: How Do Speakers Work?

Electromagnetic Pickups: A Different Transducer

A microphone reads pressure in the air; an electric guitar's **pickup** reads the string directly, and it works on a different principle entirely.

A magnetic pickup is a coil of wire wound around a permanent magnet. The magnet's field passes through the string, provided the string is ferrous (steel, or steel-wound), and magnetizes a section of it. As the string vibrates, it disturbs that field, and the changing flux through the coil induces a voltage by Faraday's law. No power supply, no diaphragm: the string *is* the diaphragm.

One consequence is easy to miss. Because the induced voltage depends on the *rate of change* of flux, the pickup's output is proportional to the string's **velocity**, not its displacement. This is the same distinction Chapter 10 draws between the bridge force a soundboard responds to and the string's actual shape.

A second consequence is audible on every electric guitar. A pickup samples the string's motion at one fixed point along its length, and Chapter 10 showed that the plucking or driving point determines which harmonics are strong: a point that is a node for some harmonic contributes almost nothing at that harmonic to the output. A **neck pickup**, positioned nearer the string's midpoint, favors the fundamental and sounds warm; a **bridge pickup**, positioned near a region where high harmonics have their antinodes, emphasizes them and sounds bright and biting. Switching pickups, or blending them, is comb filtering by another name, chosen by ear rather than calculated.

Not every electric or amplified string instrument uses this principle. A **piezoelectric pickup**, common on acoustic-electric guitars, basses, and violins, is mounted under the bridge saddle and generates a voltage from mechanical stress rather than magnetic flux, so it works on nylon or gut strings that a magnetic pickup cannot see at all, and it responds to the bridge's motion rather than the string's.

15.2 Signals and Their Imperfections

Frequency Response

The **frequency response** of any component is its gain against frequency. A perfect one would be flat across the audible range.

Nothing is. Loudspeakers are the worst offenders by a wide margin: even a good one varies by several decibels across its range, while a competent amplifier is flat to a fraction of a decibel. In any audio chain, the transducers, microphone, loudspeaker, and the room, are where essentially all of the frequency-response error lives.

Linear and Nonlinear Distortion

Two kinds of error, and conflating them causes endless confusion.

Linear distortion changes the *balance* of the frequencies present. A frequency-response error is linear distortion, and so is a phase shift. It adds nothing new: whatever comes out was in the input, at a different level. It can, in principle, be corrected by an inverse filter, which is what room equalization attempts.

Nonlinear distortion creates frequencies that were **not present in the input**. If the output is not proportional to the input, Chapter 3's superposition fails, and a sum of two tones produces components at sums and differences of their frequencies. This cannot be corrected by filtering, because the new components are entangled with the wanted ones.

Two measures are quoted. **Total harmonic distortion** measures added harmonics of a single input tone. **Intermodulation distortion** measures the sum and difference components from two tones, and is the more revealing, because harmonics of a musical note are at least musically related to it while intermodulation products are not.

> **Warning**
> Nonlinearity is not always a defect. A guitar amplifier driven into distortion is a nonlinear

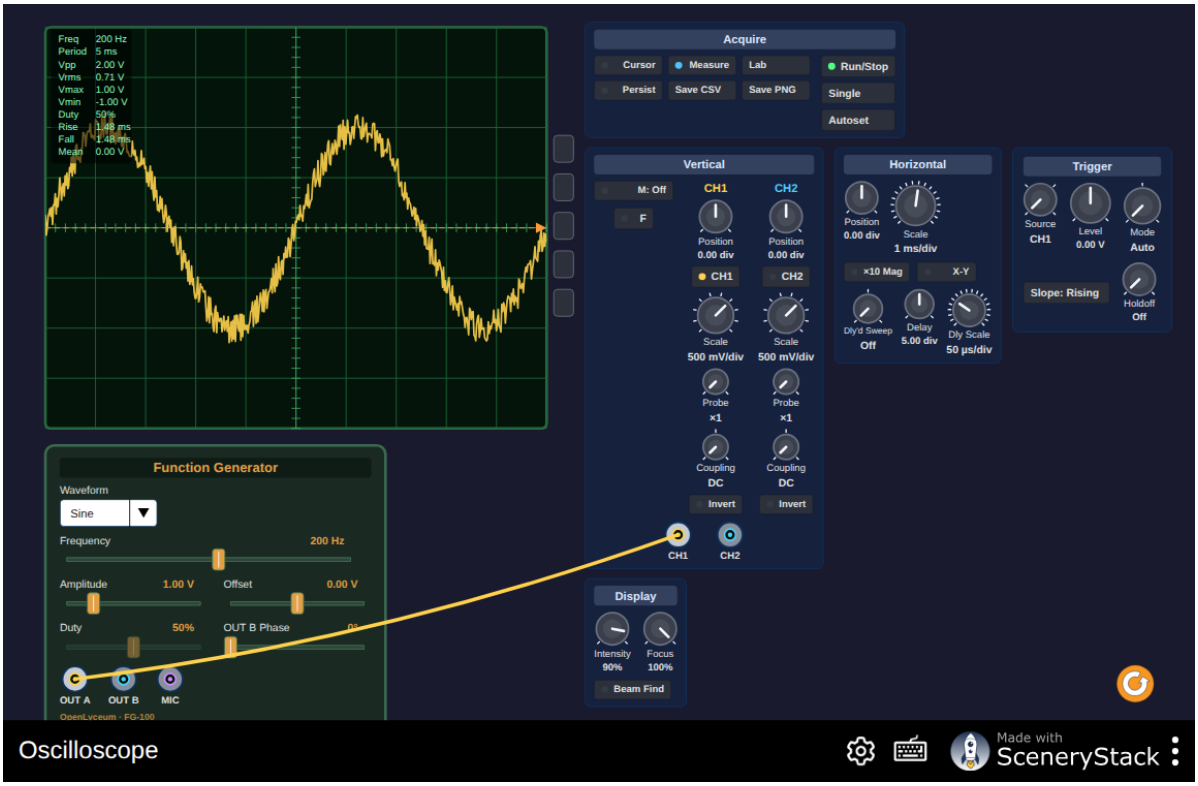


Figure 15.6: An oscilloscope with a signal generator behind it. Set a sine wave and measure its period from the screen; then add a second channel and look at what a sum looks like. This is the instrument the laboratory exercises in Laboratory Exercises assume; fluency with it helps before meeting a real one. Interactive simulation: Oscilloscope

device used deliberately, and its intermodulation products are the sound. The distinction is whether the nonlinearity is chosen. This is the same point Chapter 12 made about cymbals.

Noise and Dynamic Range

Every stage adds noise, and the **dynamic range** of a system is the span between its noise floor and the level at which it distorts.

The requirement follows from Chapter 7. An orchestra spans about 50 dB from *pianissimo* to *fortissimo*, and a good hall has a background around 25 dB SPL. A system with 90 dB of dynamic range has ample margin; one with 50 dB does not, and either the quiet passages disappear into hiss or the loud ones distort.

15.3 Digital Audio

Sampling and the Nyquist Frequency

Here is the central result.

The sampling theorem. A signal containing no frequencies above $f_{\max}$ is **completely determined** by samples taken at any rate greater than $2f_{\max}$.

Not approximately. Completely. The samples contain every bit of information in the original, and the original can be reconstructed from them exactly.

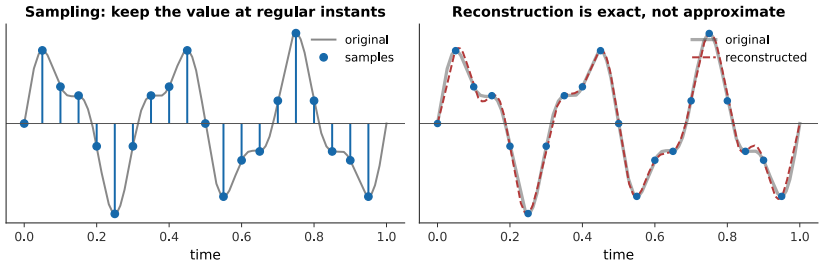


Figure 15.7: Sampling and reconstruction. The right panel is the surprising half: the reconstruction is not a join-the-dots approximation but an exact recovery, obtained by summing a sinc function centered on each sample. Nothing has been lost.

The **Nyquist frequency** is half the sampling rate: the highest frequency a given rate can represent.

The compact disc samples at 44.1 kHz, giving a Nyquist frequency of 22.05 kHz, comfortably above the 20 kHz limit of hearing. The awkward number is historical: early digital recorders stored data on video tape, and 44.1 kHz was what fitted the existing video line rates.

Aliasing and the Anti-Aliasing Filter

The theorem's condition, no frequencies above Nyquist, is not optional, and violating it is not a gentle degradation.

The consequence is severe. Aliased components are not a distortion of the original; they are *new frequencies at unrelated pitches*, and once the samples are taken there is no way to remove them.

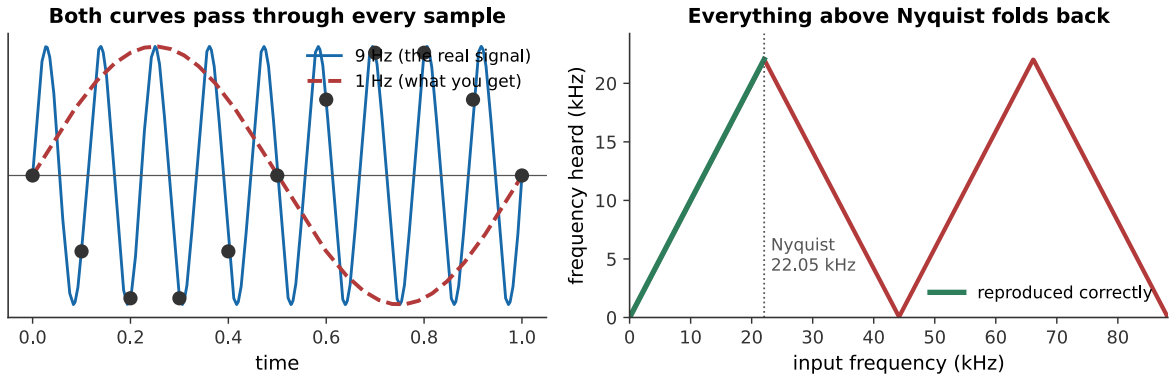


Figure 15.8: Left: a 9 Hz signal sampled at 10 Hz produces exactly the same samples as a 1 Hz signal. Nothing in the data can distinguish them. **Right**: the general rule, every frequency above Nyquist is **folded** back below it, reappearing as a different frequency altogether.

Worked example: what frequency does an alias appear at?

A synthesizer runs at a sampling rate of $f_s = 8$ kHz with no anti-aliasing filter and sweeps a tone upward through 5.5 kHz. What frequency is actually heard at that instant?

The Nyquist frequency is $f_s/2 = 4$ kHz. For an input between Nyquist and the sampling rate itself, the fold is a mirror image about Nyquist:

$$f_{\text{heard}} = f_s - f_{\text{in}} = 8000 \text{ Hz} - 5500 \text{ Hz} = 2500 \text{ Hz}. \tag{15.1}$$

Therefore, a generator climbing steadily through 5.5 kHz is heard descending through 2.5 kHz, exactly the reversal in the sweep of the audio example below. Once the samples are taken, a genuine 2.5 kHz tone and this alias of a 5.5 kHz tone are the same list of numbers; nothing in the data says which one produced it, which is why the filtering has to happen before the sampling and cannot be repaired afterward.

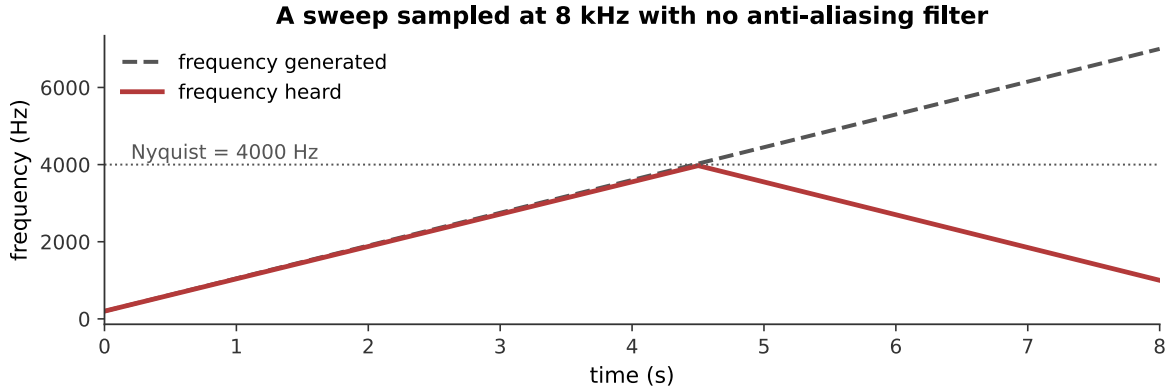


Figure 15.9: A sweep from 200 Hz to 7 kHz, synthesized at 8 kHz with no anti-aliasing filter. Above 4 kHz the tone folds back and descends. Nothing in the generator ever descends; what you are hearing is the folding diagram above, made audible. *What you hear:*A rising tone that climbs for about half the clip, then turns round and descends, and then turns again, even though the generator is sweeping steadily upward the whole time. Listen: Aliasing Sweep

The cure is to filter **before** sampling. An **anti-aliasing filter** removes everything above Nyquist while the signal is still continuous, so the theorem's condition is satisfied. It must come first, after sampling, the aliased components are indistinguishable from real ones.

This is the technical reason for sampling faster than strictly necessary. A perfect filter is impossible, so a real one needs a transition band, and the gap between 20 kHz and 22.05 kHz is where the CD's filter does its work.

Quantization, Bit Depth, and Dither

Sampling discretizes time. **Quantization** discretizes amplitude, rounding each sample to the nearest of a finite set of levels, and unlike sampling it is *not* lossless.

The error is bounded by half a level, and the resulting dynamic range is

$$\text{dynamic range} \approx 6.02\,b + 1.76 \text{ dB}, \tag{15.2}$$

with b the number of bits: the useful rule being **about 6 dB per bit**. A 16-bit CD gives 98 dB, comfortably beyond any listening room. A 24-bit studio format gives 146 dB, which is far beyond any microphone, and exists to provide headroom during processing rather than for final delivery.

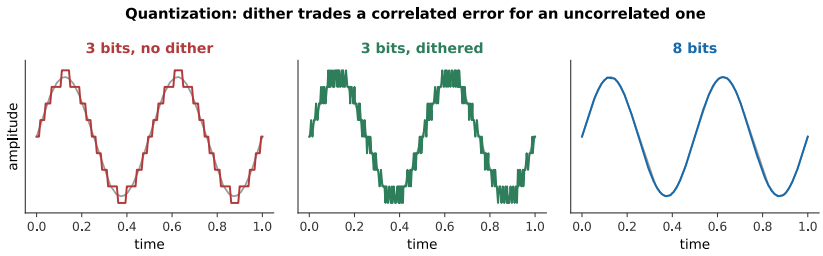


Figure 15.10: Quantization at three bit depths. The middle panel shows what **dither** does: adding a small amount of noise *before* quantizing.

Dither seems perverse, deliberately adding noise to improve quality, and it is one of the more elegant ideas in signal processing. Without it, the quantization error is **correlated** with the signal: it follows the waveform, producing harmonic distortion that is worst on quiet passages, exactly where it is least welcome. Adding a small random noise decorrelates the error, converting it into a steady, signal-independent hiss.

The total error energy is slightly greater. The audible result is much better, because the ear is far more tolerant of steady noise than of signal-dependent distortion.

15.4 Perceptual Coding

Using Masking to Throw Things Away

A CD carries about 1.4 megabits per second. An MP3 at 128 kbit/s carries a tenth of that and most listeners cannot reliably tell them apart. Something is being thrown away, and the question is what.

The answer comes straight from Chapter 7. **If a component is masked, the listener cannot hear it; so it does not need to be stored.**

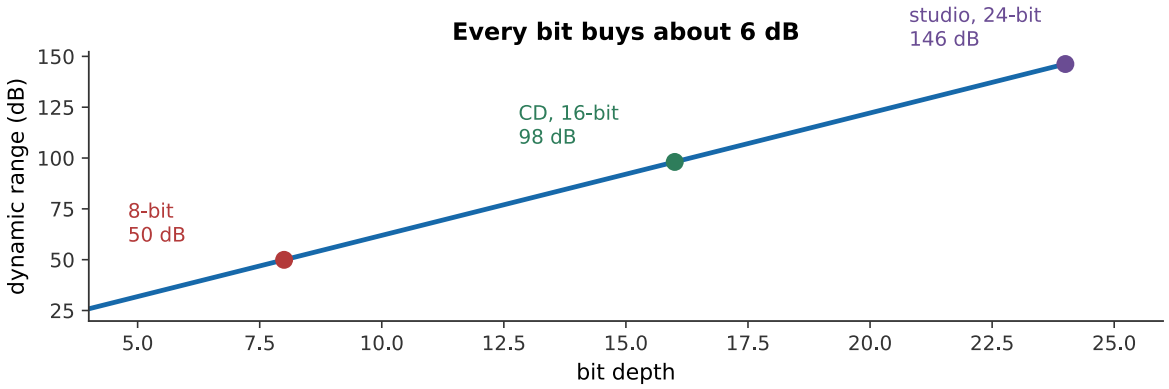


Figure 15.11: The same note at three quantizations. Listen to the **tails**. The undithered version does not merely get noisy as it decays, it becomes distorted and granular, because the error is following the signal. The dithered version has more noise overall and sounds better, which is the whole argument for dither. *What you hear:*The same decaying note three times. The first is clean. The second is clean at first and then breaks up into a gritty, buzzy crackle as it fades. The third has an audible steady hiss but fades away smoothly. Listen: 16-bit, 6-bit undithered, 6-bit dithered

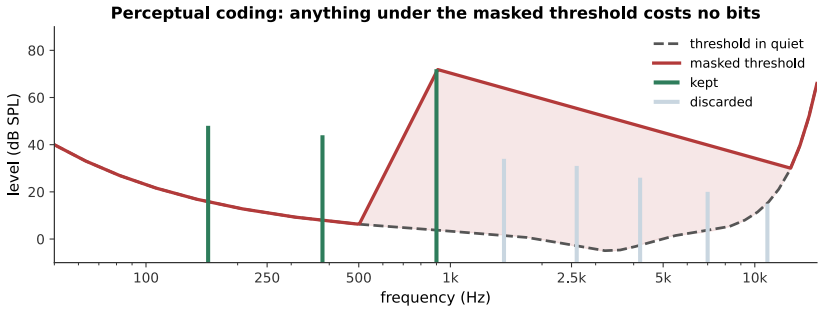


Figure 15.12: How a codec decides. It computes the **masked threshold**, the level below which nothing is audible, given what else is present, and discards everything under it. The discarded components are not approximated or attenuated; they are simply not stored.

A codec works in short blocks, and for each one it transforms to the frequency domain, computes the masked threshold from a psychoacoustic model, allocates bits so that quantization noise in each band sits just below that threshold, and encodes the result.

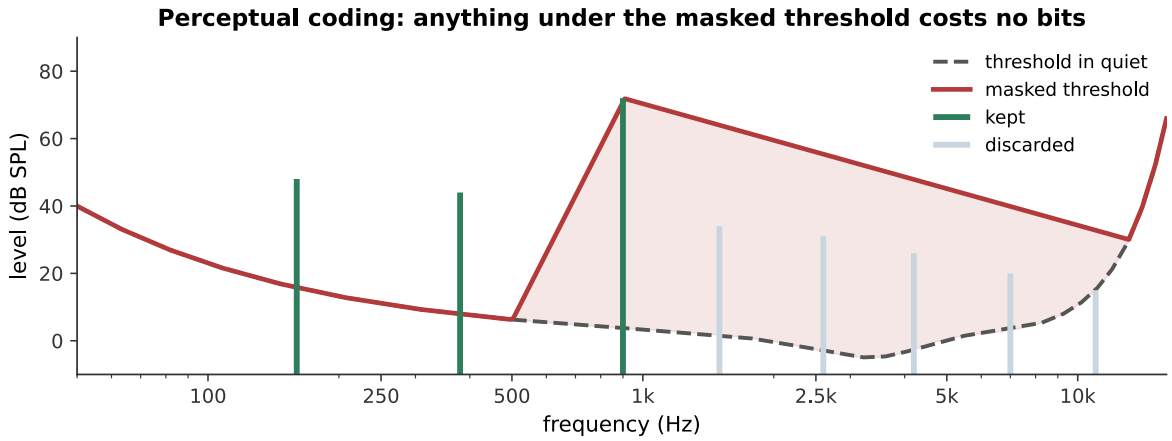


Figure 15.13: A loud 900 Hz tone with three quiet partials, and the same with those partials deleted. They are hard to tell apart because the partials were masked, they were present in the signal and absent from your perception of it, so removing them changed nothing you could hear. *What you hear:*Two tones that sound essentially identical. The second has had three quiet partials removed entirely, and it is very hard to hear any difference at all. Listen: All components, Masked components removed

What a Lossy Codec Keeps

The bit allocation follows the ear's own priorities, and the consequences are visible in what codecs are good and bad at.

Low frequencies get more bits, because the critical bands are narrower there and masking is less generous.

Steady tones compress well; sharp transients compress badly. A transient needs short blocks to be located in time, and short blocks have poor frequency resolution (Chapter 5's trade-off, in a new guise). Applause, castanets, and harpsichords are the classic hard cases, and codecs switch block length dynamically for that reason.

Stereo is exploited. Much of the two channels is common, and encoding a sum and a difference is cheaper than encoding both.

What Is Lost for Good

Two honest caveats.

Lossy is lossy. The discarded information is gone. Re-encoding an MP3, decoding and re-encoding, as happens when a file is edited or converted, applies the model to a signal that has already been altered, and the errors accumulate. Archives should be stored losslessly.

The model can be wrong. The masking model is built on average listeners and typical signals. Unusual material can defeat it, and the artifacts are characteristic: pre-echo, where a transient's quantization noise is spread backwards in time and becomes audible *before* the event that caused it, and a swirling, watery quality in cymbals where the codec's decisions change from block to block.

15.5 Synthesis

Additive Synthesis: Fourier in Reverse

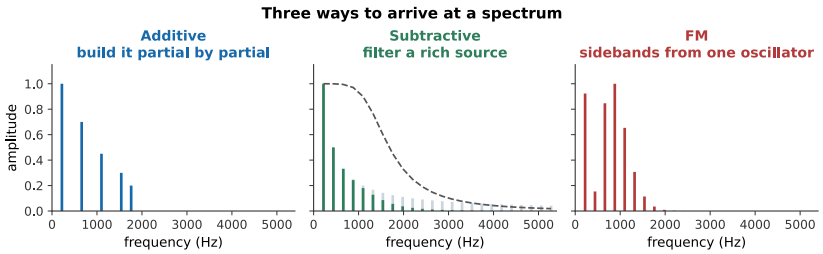


Figure 15.14: Three routes to a spectrum. Additive builds it partial by partial, subtractive carves it out of a rich source, and FM generates a whole family of sidebands from two oscillators.

Additive synthesis builds a sound by summing sinusoids with individually chosen amplitudes and envelopes. It is Fourier's theorem run backwards, and it can produce any periodic sound whatever.

Its weakness is not sound quality but **control**. A realistic instrument needs dozens of partials, each with its own time-varying amplitude, which is hundreds of parameters per note. The Hammond organ is additive synthesis with nine drawbars, and it sounds like a Hammond organ rather than like anything else, which says something about how far nine partials get you.

Subtractive Synthesis and the Source–Filter Model Again

Subtractive synthesis starts with a spectrally rich source, a sawtooth, a square, or noise, and removes what is not wanted with a filter.

This is exactly the **source–filter model** of Chapter 13, and the analogy is not loose: the oscillator is the vocal folds and the filter is the vocal tract. A resonant low-pass filter with a swept cutoff is imitating a formant, and a synthesizer player sweeping a filter is doing what a singer does moving their tongue.

It dominated analog synthesis because it is efficient. Two or three controls, cutoff, resonance, and an envelope on the cutoff, produce an enormous range of usable timbres, where additive synthesis needs hundreds.

FM Synthesis and Sampling

FM synthesis, discovered by John Chowning in 1967 and commercialized in the Yamaha DX7 in 1983, uses one oscillator to modulate another's frequency.

Modulating a carrier at f_c with a modulator at f_m produces sidebands at

$$f_c \pm n f_m, \qquad n = 0, 1, 2, \ldots \tag{15.3}$$

whose amplitudes are Bessel functions of the **modulation index**: the ratio of frequency deviation to modulator frequency. Raising the index spreads energy into more sidebands.

Two features made it commercially decisive. It is **cheap**: two oscillators produce dozens of partials, which mattered enormously with 1980s hardware. And the **index can be varied over the note**, so a falling index makes a bright attack settle to a pure sustain, precisely

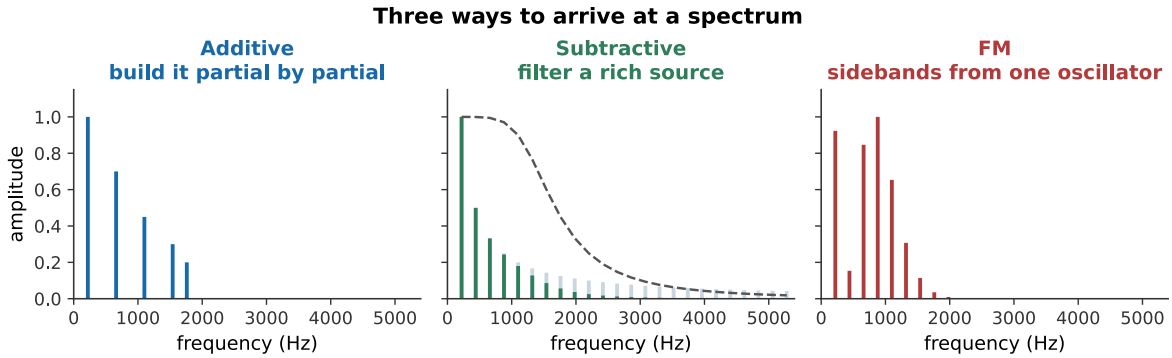


Figure 15.15: The same pitch by three methods. The subtractive example's filter sweep and the FM example's falling modulation index are doing the same job, making the spectrum change over the course of the note, which Chapter 5 identified as the thing static synthesis most obviously lacks. *What you hear:*Three tones at the same pitch. The first is hollow and organ-like. The second starts bright and darkens as its filter closes. The third is metallic and bell-like, becoming purer as it decays. Listen: Additive, Subtractive, FM

the behavior of a struck bar or a bell (Chapter 12). FM's metallic and bell-like sounds were therefore much better than anything else available at the time.

If f_c and f_m are in a simple ratio, the sidebands land on a harmonic series and the sound is pitched. If not, they do not, and the result is inharmonic, which is the same distinction Chapter 12 drew between a string and a gong, arrived at from the other direction.

Sampling, playing back recordings of real instruments, sidesteps synthesis entirely and dominates commercial music production. Its difficulty is that a sample is one note, at one dynamic, at one moment; covering an instrument's range requires many, and the couplings Chapter 10 described: a piano played louder is also *brighter*, must be reproduced by switching between samples recorded at different dynamics. A sampler that changes only loudness sounds instantly wrong.

15.6 Space in a Recording

Two Speakers, One Image

Two loudspeakers can produce a sound that appears to come from a point between them, where nothing is.

The mechanism is **summing localization**. Identical signals from both give a phantom image at the center. Making one louder: an **intensity** difference, pulls the image toward it, and about 15–18 dB places it fully at one speaker. Making one earlier: a **time** difference, does the same, with about 1 ms sufficing.

Both work because they crudely reproduce what a real source would do at the two ears.

Localization Cues and the HRTF

Real localization uses three cues, and a recording can only approximate them.

Interaural time difference dominates below about 1.5 kHz: a sound from the left reaches the left ear up to 0.7 ms sooner.

Interaural level difference dominates above: the head shadows the far ear, by up to 20 dB at high frequencies.

Spectral cues from the pinna resolve front from back and above from below (Chapter 6), and they are the only cue that can. The whole direction-dependent filter is called the **head-related transfer function**.

Ordinary stereo reproduces only the first two, and only crudely, so a stereo image is a line between the speakers rather than a space. **Binaural** recording, made with microphones in the ears of a dummy head, captures the HRTF as well, and over headphones produces a startling sense of height and behindness. Its limitation is that it encodes *somebody else's* ears, and that it breaks as soon as the listener turns their head, hence the rise of head-tracked systems.

Reverberation, Real and Artificial

Every recording must decide how much room to include, and the decision cannot be postponed: reverberation recorded into a track cannot be removed.

Modern practice therefore records **dry**, close-miked, in a treated room, and adds reverberation afterwards. Two methods.

Algorithmic reverberation builds a decaying field from networks of delays and filters. It is efficient, adjustable in every parameter, and can produce spaces that could not be built.

Convolution reverberation uses a measured impulse response of a real hall and convolves the dry signal with it (Chapter 14). It sounds exactly like that hall, from that seat, because it *is* that hall's response, at the cost of being fixed, since the response encodes one source position, one listening position, and one room.

And so the book's last idea is one of its first. Chapter 3 established superposition; Chapter 14 established that a room's effect is completely described by its impulse response; and convolution reverb is nothing but those two facts applied at forty-four thousand samples per second. A recording engineer placing a soloist in the Concertgebouw is doing linear superposition, very fast.

15.7 Summary

- **Dynamic microphones** use a coil and magnet; **condensers** use a variable capacitor and a much lighter diaphragm. Both want a well-damped resonance outside the audio band, low Q, as in Chapter 4.
- **Polar patterns** follow from whether the diaphragm senses pressure (omnidirectional) or pressure difference (figure-of-eight); cardioid is their sum. Every pressure-difference microphone shows **proximity effect**.
- **Loudspeakers need multiple drivers**, because directivity depends on size against wavelength and music spans a factor of a thousand in wavelength. Enclosures prevent front-to-back cancellation; a bass-reflex port is a Helmholtz resonator.
- **A magnetic pickup** induces a voltage from a ferrous string's motion directly, no diaphragm required; its output tracks string velocity, and its fixed sampling point along the string means placement (neck versus bridge) reshapes the harmonic balance, the same driving-point physics as Chapter 10.
- **Linear distortion** rebalances existing frequencies and is in principle correctable; **nonlinear distortion** creates new ones and is not.
- **The sampling theorem**: a signal with no energy above $f_{\max}$ is *completely* determined by samples at any rate above $2f_{\max}$. Reconstruction is exact.

- **Aliasing** folds everything above the Nyquist frequency back below it, producing new frequencies at unrelated pitches. The anti-aliasing filter must act **before** sampling, because afterwards nothing can distinguish an alias from a real component.
- **Quantization gives about 6 dB of dynamic range per bit. Dither** adds noise before quantizing, converting signal-correlated distortion into steady hiss, more total error energy, much better sound.
- **Perceptual coding discards masked components**, which by definition cannot be heard. It struggles with transients, because locating them in time requires short blocks and short blocks resolve frequency poorly.
- **Additive synthesis** specifies every partial; **subtractive** filters a rich source and is the source–filter model again; **FM** generates Bessel-weighted sidebands from two oscillators, giving inharmonic spectra cheaply.
- **Stereo works by summing localization**, reproducing interaural time and level differences but not the pinna cues that distinguish front from back. **Convolution reverb** is superposition plus a measured impulse response.

15.8 Check Your Understanding

Five short, auto-graded questions cycle within one compact activity, using five different response styles.

1. **Multiple choice.** Where must an anti-aliasing low-pass filter act: before sampling, after sampling, or only during playback?

2. **True or false.** Dither improves low-level digital sound by reducing the total error energy introduced during quantization.

3. **Drag the words.** Match dynamic and condenser microphones and omnidirectional and figure-of-eight patterns to their operating principles.

4. **Fill in the blank.** At roughly 6 dB per bit, ideal 16-bit audio provides about _____ dB of dynamic range.

5. **Mark the words.** Identify the two processes that create new frequency components: "FM synthesis generates sidebands, and nonlinear distortion creates new frequencies; subtractive filtering only rebalances frequencies already present."

15.9 Conceptual Questions

1. Explain why a condenser microphone tracks transients better than a dynamic one.

2. Explain why an omnidirectional microphone shows no proximity effect while a cardioid does.

3. Explain why a single loudspeaker driver cannot reproduce the whole audible range, referring to source size and wavelength.

4. A guitarist blends a neck and a bridge pickup and notices the tone changes even though nothing about the string or the amplifier has changed. Explain why, referring to the driving-point argument of Chapter 10.

5. Distinguish linear from nonlinear distortion, and explain why only one of them can in principle be corrected.

6. State the sampling theorem, and explain what is surprising about the word “completely” in it.

7. Explain why an anti-aliasing filter must be applied before sampling rather than after.

8. Explain why adding noise before quantizing improves the perceived quality, given that it increases the total error.

9. Explain why perceptual codecs handle a sustained organ chord better than applause.

15.10 Problems

Exercise 15.1
(Straightforward) A system samples at 48 kHz. (a) What is the Nyquist frequency? (b) What is the highest frequency it can represent? (c) A 30 kHz component reaches the converter unfiltered. What frequency appears?

Solution to Exercise 15.1
(a) $48000/2 = 24$ kHz.
(b) Anything below 24 kHz.
(c) Frequencies above Nyquist fold back:

$$f_{\text{alias}} = |48000 - 30000| = 18 \text{ kHz}. \tag{15.4}$$

Therefore a 30 kHz component, inaudible in itself, appears as an 18 kHz component, which is audible to many listeners and is not in the original at all.

Exercise 15.2
(Straightforward) A CD samples at 44.1 kHz with 16 bits per sample, in stereo. (a) What is the data rate? (b) How many megabytes is a 74-minute disc?

Solution to Exercise 15.2
(a) Samples per second times bits per sample times channels:

$$44100 \times 16 \times 2 = 1.41 \times 10^{6} \text{ bits/s} = 1.41 \text{ Mbit/s}. \tag{15.5}$$

(b) 74 minutes is 4440 s:

$$1.41 \times 10^{6} \times 4440 = 6.27 \times 10^{9} \text{ bits} = 7.83 \times 10^{8} \text{ bytes} = 783 \text{ MB}. \tag{15.6}$$

Therefore about 780 MB, which is the familiar capacity of an audio CD: the format’s storage was designed around exactly this calculation.

Exercise 15.3
(Moderate) Find the dynamic range of (a) an 8-bit system, (b) a 16-bit system, (c) a 24-bit system. (d) Which is adequate for a concert hall with a 25 dB background and a 105 dB peak?

Solution to Exercise 15.3

Using $6.02b + 1.76$ dB:

(a) $6.02(8) + 1.76 = 49.9$ dB.

(b) $6.02(16) + 1.76 = 98.1$ dB.

(c) $6.02(24) + 1.76 = 146.2$ dB.

(d) The hall requires $105 - 25 = 80$ dB. The 8-bit system is inadequate; 16-bit has 18 dB of margin and is sufficient; 24-bit is far beyond any microphone's own noise floor and exists to give headroom during processing rather than for delivery.

Exercise 15.4

(Straightforward) An MP3 at 128 kbit/s replaces a CD stream. (a) What is the compression ratio? (b) What fraction of the original data remains? (c) What justifies discarding the rest?

Solution to Exercise 15.4

(a) $1411/128 = 11.0$.

(b) $1/11 = 9.1\%$.

(c) **Masking** (Chapter 7). At any instant, most of the spectral detail in a musical signal lies below the masked threshold set by the louder components, and a listener cannot hear it. A codec computes that threshold and spends bits only above it.

Therefore the discarded 91% is not "detail the listener might miss"; it is information the auditory system was never going to receive. The compression is therefore much more effective than a general-purpose lossless compressor, which manages only about 2:1 on audio.

Exercise 15.5

(Moderate) An FM synthesizer has a carrier at 440 Hz and a modulator at 440 Hz with index 3. (a) Where are the first four sideband pairs? (b) Is the result harmonic? (c) Repeat with a modulator at 311 Hz.

Solution to Exercise 15.5

(a) Sidebands are at $f_c \pm n f_m$:

$$440{\pm}440 = 880, 0; \quad 440{\pm}880 = 1320, -440; \quad 440{\pm}1320 = 1760, -880; \quad 440{\pm}1760 = 2200, -1320. \tag{15.7}$$

Negative frequencies fold back with inverted phase, so the set is 440, 880, 1320, 1760, 2200 Hz.

(b) **Yes**, every component is a multiple of 440 Hz, so the result is a harmonic series and the sound is clearly pitched.

(c) With $f_m = 311$ Hz: $440 \pm 311 = 751, 129$; $440 \pm 622 = 1062, -182$; $440 \pm 933 = 1373, -493$. The set is 129, 182, 440, 493, 751, 1062, 1373 Hz, no common fundamental, so the result is **inharmonic** and bell-like.

Therefore the ratio $f_c : f_m$ decides whether FM produces a pitched instrument or a percussion instrument, which is exactly the distinction of Chapter 12 reached from the other direction.

Exercise 15.6

(Moderate) A stereo pair is fed a signal with the left channel 6 dB louder. (a) Where does the phantom image appear? (b) What time difference would place it similarly? (c) Why do both work?

Solution to Exercise 15.6

(a) Full deflection to one speaker takes about 15–18 dB, so 6 dB places the image roughly a third of the way from center to the left speaker.

(b) Full deflection takes about 1 ms, so a comparable shift needs roughly 0.3 ms of delay on the right channel.

(c) Both crudely reproduce what a real source to the left would produce at the ears: it would arrive sooner and louder at the left ear. The auditory system integrates the two loudspeakers' outputs and infers a direction consistent with the cues it receives.

Therefore intensity and time panning are interchangeable to a first approximation, and are used interchangeably, though they are not identical, and combining them inconsistently produces an unstable image.

Exercise 15.7

(Moderate) A loudspeaker's woofer is 25 cm across and its tweeter 25 mm. (a) Above what frequency does each become larger than a wavelength? (b) Suggest a crossover frequency. (c) Explain the reasoning.

Solution to Exercise 15.7

(a) Setting $\lambda = d$, so $f = v/d$:

$$f_{\text{woofer}} = \frac{343 \text{ m/s}}{0.25 \text{ m}} = 1.37 \text{ kHz}, \qquad f_{\text{tweeter}} = \frac{343 \text{ m/s}}{0.025 \text{ m}} = 13.7 \text{ kHz}. \tag{15.8}$$

(b) Somewhere around 1.5–2.5 kHz.

(c) The crossover should hand over before the woofer starts beaming: that is, at or below about 1.4 kHz by the calculation above, though in practice a little above is tolerated. It must not be so low that the tweeter is asked to move air at frequencies where its small cone cannot produce enough displacement without distorting.

Therefore the choice is squeezed from both sides, and 2 kHz is the usual compromise, which is unfortunate, since it sits in the region where hearing is most sensitive and where the ear most easily detects the discontinuity.

Exercise 15.8

(Moderate) Explain quantitatively why a 44.1 kHz sampling rate is used rather than exactly 40 kHz, given that hearing extends to 20 kHz.

Solution to Exercise 15.8

At exactly 40 kHz the Nyquist frequency is 20 kHz, so the anti-aliasing filter would need to pass 20 kHz unattenuated and stop everything above 20 kHz completely: an infinitely steep transition, which is impossible.

At 44.1 kHz the Nyquist frequency is 22.05 kHz, leaving a transition band from 20 to 22.05 kHz. Expressed as a slope requirement, the filter must fall from passband to stopband over

$$1200 \log_2\left(\frac{22050}{20000}\right) = 169 \text{ cents}, \tag{15.9}$$

under a tone and a half, still a demanding filter, but a buildable one.

Therefore the extra 4.1 kHz buys the filter room to work. The precise figure is historical rather than acoustic: early digital audio was stored on video recorders, and 44.1 kHz is what fell out of the available line and field rates.

Exercise 15.9
(Moderate) A guitar amplifier is driven into distortion. Two notes at 220 Hz and 330 Hz are played. (a) Name four intermodulation products. (b) Are any of them musically related to the input? (c) Comment on why this is nonetheless a popular sound.

Solution to Exercise 15.9
(a) Sums and differences and their combinations:

$$330 - 220 = 110 \text{ Hz}, \quad 330 + 220 = 550 \text{ Hz}, \tag{15.10}$$

$$2(220) - 330 = 110 \text{ Hz}, \quad 2(330) - 220 = 440 \text{ Hz}. \tag{15.11}$$

(b) Yes, in this case, all of them. The two inputs are a perfect fifth apart (3:2), so they are harmonics 2 and 3 of a common fundamental at 110 Hz, and every intermodulation product is therefore also a harmonic of 110 Hz.

(c) That is precisely why the sound works. For **consonant** intervals the distortion products fall on the existing harmonic series and add weight rather than dissonance. Power chords, built on fifths, sound so good through a distorted amplifier for that reason.

For a more complex chord, a major third, say, the products do *not* all fall on a common series, and the result is muddy. Guitarists know this as the rule that distorted chords should be kept to two notes, and the reason is the arithmetic above.

Exercise 15.10
(Challenging) A convolution reverb uses an impulse response 3.5 s long, sampled at 48 kHz. (a) How many samples? (b) A naive convolution of a 4-minute track requires how many multiply operations? (c) Comment.

Solution to Exercise 15.10
(a) $3.5 \times 48000 = 1.68 \times 10^5$ samples.

(b) A 4-minute track is $240 \times 48000 = 1.15 \times 10^7$ samples, and each output sample requires one multiply per impulse-response sample:

$$1.15 \times 10^7 \times 1.68 \times 10^5 = 1.9 \times 10^{12} \text{ multiplies}. \tag{15.12}$$

(c) Two trillion multiplications per channel is far too many for real time on modest hardware.

The remedy is the **convolution theorem**: convolution in time is multiplication in frequency, so the operation can be done by transforming both signals, multiplying, and transforming back. Using the fast Fourier transform this reduces the cost by roughly a factor of $N / \log_2 N$, which here is about 10^4, bringing two trillion operations down to a few hundred million, which is entirely practical.

Therefore convolution reverb is possible because of Fourier's theorem, which is a pleasing place for a book that began with a vibrating string to end.

Exercise 15.11
(Challenging) Show that quantization gives approximately 6 dB of dynamic range per bit.

Solution to Exercise 15.11

With b bits there are 2^b levels. Let the full-scale signal be a sine of amplitude A, so peak-to-peak it spans $2A$ and each quantization step is

$$\Delta = \frac{2A}{2^b}. \tag{15.13}$$

The quantization error is uniformly distributed over $\pm\Delta/2$, so its mean square is

$$\langle e^2 \rangle = \frac{\Delta^2}{12}. \tag{15.14}$$

The signal's mean square is $A^2/2$. The signal-to-noise ratio is therefore

$$\frac{A^2/2}{\Delta^2/12} = \frac{A^2/2}{(2A/2^b)^2/12} = \frac{12\,A^2\,2^{2b}}{2 \cdot 4A^2} = \frac{3}{2}\,2^{2b}. \tag{15.15}$$

In decibels:

$$10\log_{10}\left(\frac{3}{2}\,2^{2b}\right) = 10\log_{10}(1.5) + 20b\log_{10}(2) = 1.76 + 6.02\,b \text{ dB}. \tag{15.16}$$

Therefore each additional bit doubles the number of levels, halves the step, and buys $20\log_{10}2 = 6.02$ dB, and the 1.76 dB offset comes from the ratio between a sine's mean square and a uniform error's.

Exercise 15.12

(Straightforward) A recording is made with a cardioid microphone 30 cm from a singer, in a room with a critical distance of 1.2 m. (a) Is the recording dominated by direct or reverberant sound? (b) The engineer moves to 2.5 m. What changes? (c) Which would be preferred, and why?

Solution to Exercise 15.12

(a) At 30 cm the microphone is well inside the critical distance, so the direct sound dominates heavily, this is a dry, close recording.

(b) At 2.5 m the microphone is beyond the critical distance, so the reverberant field dominates and the recording captures mostly room.

(c) For most modern production, the **close** position. A dry recording can have reverberation added afterwards by convolution (Chapter 14), and the amount can be chosen at mixing time. Reverberation recorded into a track cannot be removed.

Two caveats. The close position will show **proximity effect**, boosting the bass substantially, which must be corrected or exploited. And if the room is genuinely good, a fine hall, the distant position captures something an artificial reverberation cannot quite reproduce. Classical recording still uses distant main pairs for that reason.

Exercise 15.13

(Moderate) Explain, using Chapter 5's time–frequency trade-off, why perceptual codecs produce "pre-echo" on sharp transients.

Solution to Exercise 15.13

A codec works on blocks. Within each block it transforms to the frequency domain, decides how much quantization noise each band can carry without being audible, and quantizes accordingly.

The quantization noise it introduces is spread **uniformly over the whole block**, because the quantization happens in the frequency domain and a frequency-domain error has no localization in time.

Now consider a block containing a castanet click: silence for most of its duration, then a loud transient. The masking model computes a threshold from the block's overall spectrum, which the loud transient dominates, so a generous amount of noise is permitted. That noise is then spread across the entire block, including the silent part *before* the click.

The result is a burst of noise audible in the silence preceding the transient: **pre-echo**.

Chapter 5's trade-off is what makes this unavoidable in principle: locating the transient precisely in time requires a short block, and a short block resolves frequency poorly, which degrades the masking model. Codecs respond by **switching block length dynamically**, using short blocks when a transient is detected and long ones otherwise, which mitigates the problem without eliminating it.

Exercise 15.14

(Moderate) A sampler reproduces a piano by pitch-shifting one recorded note. (a) What goes wrong if a note recorded at C4 is shifted up an octave? (b) What goes wrong if only one dynamic level was recorded? (c) Relate both to Chapter 10.

Solution to Exercise 15.14

(a) Pitch-shifting by resampling scales *every* frequency in the recording, including the resonances of the soundboard and body. But those resonances are properties of the instrument and do **not** move when the pitch changes, they are formants, in the sense of Chapter 13. Shifting an octave moves them an octave too, and the result sounds like a smaller instrument. Shifting down produces the well-known "munchkinization" in reverse.

The note's *duration* also scales, so the shifted note decays twice as fast, which is wrong, since a real piano's high notes decay faster for reasons of string mass, not by a factor of exactly two.

(b) Chapter 10 showed that a piano hammer's felt is nonlinear: struck harder, contact time shortens and more high harmonics survive. **A piano played louder is brighter, not merely louder.** A sampler with one dynamic level can only change the volume, so a *fortissimo* sounds like a *pianissimo* turned up, instantly recognizable as wrong.

(c) Both failures come from the same source: an instrument is not a single sound with adjustable pitch and volume. Its spectrum depends on the note and on the dynamic, through mechanisms Chapter 10 described.

Therefore good samplers record many notes across the range, at several dynamic levels, and cross-fade between them. A convincing sampled piano therefore runs to many gigabytes.

Part VII

Appendices

Chapter 16

Teaching Guide

This guide is for an instructor putting the book on a fifteen-week calendar. The chapters are titled by subject rather than by week, so that the book survives a change of syllabus, but they were sized for one week each and the mapping below is the straightforward one.

16.1 Suggested Fifteen-Week Pacing

Week	Assign	Purpose
1	Chapter 1, all sections	Establish what sound is, and get simple harmonic motion in hand as the model vibration.
2	Chapter 2, all sections	Turn a local vibration into a traveling wave; fix $v = f\lambda$ and the speed of sound.
3	Chapter 3, all sections	Superposition, then the standing wave: the single most important idea in the book.
4	Chapter 4, all sections	Resonance and normal modes; why an instrument responds to some frequencies and not others.
5	Chapter 5, all sections	Fourier analysis and timbre. The hinge of the course: everything after this uses spectra.
6	Chapter 6, all sections	The ear as a mechanical and neural instrument.
7	Chapter 7, all sections	Decibels and loudness. The one week with sustained arithmetic; budget problem time.
8	Chapter 8, all sections	Pitch, beats, and where consonance comes from: the bridge into tuning.
9	Chapter 9, all sections	Tuning systems. Follows directly from Week 8 and should not be separated from it.
10	Chapter 10, all sections	Strings: the family whose physics the course has already built.
11	Chapter 11, all sections	Winds and air columns.
12	Chapter 12, all sections	Percussion, and the first inharmonic spectra in the book.
13	Chapter 13, all sections	The voice, as an application of source–filter thinking.
14	Chapter 14, all sections	Room acoustics; what happens to the sound after it leaves the instrument.
15	Chapter 15, all sections; project presentations	Recording and synthesis, and the final projects.

16.2 Dependencies

Most chapters depend only on what came before, but four dependencies are strong enough to be worth naming, because breaking them costs more than it saves:

- **Chapter 5 depends on Chapter 3.** Fourier analysis is standing waves counted up; a class that has not seen the harmonic series of a string will take the spectrum on faith.
- **Chapter 9 depends on Chapter 8.** Tuning systems are an answer to a question , why do simple ratios sound good?, that Chapter 8 asks. Taught first, they are arithmetic without motivation.
- **Chapters 10–13 all depend on Chapter 5.** Every instrument chapter is an argument about a spectrum.
- **Chapter 14 depends on Chapter 7.** Reverberation time is a statement about decay in decibels.

Chapters 6 and 7 can be swapped. Chapter 2's Doppler section is used nowhere else and can be cut without consequence.

16.3 A Ten-Week Version

If the term is shorter, the following keeps the spine intact:

Week	Assign
1	Chapter 1; Chapter 2, §§2.1–2.3
2	Chapter 3, all sections
3	Chapter 4, §§4.1–4.3, 4.5
4	Chapter 5, all sections
5	Chapter 6, §§6.2–6.4; Chapter 7, §§7.1–7.3
6	Chapter 7, §§7.4–7.5; Chapter 8, §§8.1–8.3
7	Chapter 8, §8.4; Chapter 9, all sections
8	Chapter 10, all sections
9	Chapter 11, §§11.1–11.5; Chapter 12, §§12.1–12.3
10	Chapter 13, §§13.1–13.3; Chapter 14, §§14.1–14.2

16.4 What To Trim First

If the term gets tight, the best places to compress, in order:

- §2.5, the Doppler effect. Standard, enjoyable, and used nowhere later.
- §9.4, meantone and the well temperaments. The Pythagorean comma and equal temperament carry the argument; the historical middle can be a reading.
- §12.4–12.5, plates, bells, and tuning the inharmonic. Chapter 12 makes its point with bars and membranes alone.
- §15.4–15.6. Chapter 15 works as a two-session survey of microphones and digital audio.
- Chapter 6's §6.1 and §6.5. The impedance-matching argument in §6.2 and the place principle in §6.3–6.4 are what Chapters 7 and 8 need.

Do not trim Chapter 5, and do not trim Chapter 3's treatment of standing waves. Everything else in the book leans on them.

16.5 The Laboratory and the Project

Laboratory Exercises holds measurements that can be run as scheduled lab sessions or assigned as take-home work with a phone. Each is keyed to the week it supports; the natural schedule runs them in Weeks 2, 4, 5, 7, 8, 9, and 11.

The Final Project is designed to be assigned in Week 3 or 4, with a proposal due around Week 7 and presentations in Week 15. It carries the course's communication outcome, and it needs that much lead time because the hardest part, finding a question that can actually be measured, is the part students most reliably underestimate.

16.6 On Assessment

The problem sets are the obvious assessment, but two of the stated learning outcomes are not reachable through them. Collecting and interpreting experimental data is assessed by the laboratory write-ups; explaining physics to a non-specialist is assessed by the project presentation, and it is worth saying explicitly in the rubric that a technically correct talk which loses its audience has not met the outcome.

Chapter 17

Laboratory Exercises

The claims this book makes about spectra, air columns, beats, and reverberation are all measurable, most of them in an afternoon, several of them with a telephone. This appendix is a set of ten exercises that take the book's assertions and check them.

They are written to be run with whatever is available. Where a laboratory has an oscilloscope and a function generator, use them: reading a waveform off a calibrated screen builds an intuition that a software display does not. Where it does not, every exercise here has a software path, and that path is now a browser tab. The simulations named below open from a link, install nothing, ask for no account, and read the laptop's own microphone; the only downloads this appendix still asks for are for the two exercises that need to *edit* a recording.

17.1 Equipment

The short list. A laptop, a pair of headphones, and a phone will do all ten exercises at reduced precision.

Item	Laboratory version	Substitute
Signal source	Function generator	The generator built into the Oscilloscope simulation: sine, square, triangle, sawtooth, pulse, and noise, with amplitude, `DC` offset, and duty cycle. It *draws* its output rather than playing it, so for a tone you have to hear, use the Composer screen of Wave Composer
Waveform display	Oscilloscope	The same Oscilloscope simulation, driven by its own generator or by the microphone: two channels, `AC`/`DC`/`GND` coupling, triggering, `X-Y` and `FFT` modes, Δt and ΔV cursors, live frequency, period, V_{pp} and V_{rms} readouts, and `CSV` export
Spectrum analyzer	Dedicated analyzer, or a scope with `FFT`	The Oscilloscope's `FFT` mode for one spectrum at a time; the Analyzer screen of Wave Composer for a live spectrogram with the instantaneous spectrum and an `LPC` envelope beside it. Audacity's *Analyze >Plot Spectrum* or Sonic Visualiser if you would rather work from a saved file
Sound level meter	Calibrated Type 2 meter	Nothing in a browser substitutes for this one. A phone `SPL` app, **uncalibrated**, so use it for *differences*, never absolute levels
Microphone	Measurement microphone	The laptop's built-in microphone, which both simulations read directly once you grant the browser permission
Miscellaneous	Meter rule, thermometer, tuning fork, a length of string, a set of masses, a pulley, a long tube or a graduated cylinder	A tape measure, an outdoor thermometer or the day's weather report, a tuning-fork app, fishing line, kitchen scales and any objects you can hang from them, a smooth rod or a pencil taped to a table edge for the pulley, and a cardboard tube or a tall bottle for the air column. Where the apparatus is missing outright, What a Simulation Can and Cannot Replace says which exercise a

One warning about phone apps. A phone microphone rolls off sharply below about 100 Hz and its automatic gain control fights you. Turn off any noise-suppression or "voice enhancement" setting before recording, and never trust a phone for an absolute sound level. The same warning applies to a laptop microphone feeding a simulation: what you are measuring is the *shape* of a spectrum and the *ratios* of frequencies, never an absolute level.

What a Simulation Can and Cannot Replace

Three of the book's simulations do most of the bench work. The Oscilloscope is a function generator and a dual-channel scope in one window; Wave Composer synthesizes tones and analyzes live or recorded sound; and Standing Waves and Resonance supply the apparatus of two exercises whose hardware is the hardest to improvise.

Where a simulation appears in the table below, it is genuinely the instrument, not a picture of one: you set the controls, read the numbers off the screen, and write them up the same way. Where the entry says *rehearsal*, the quantity the exercise asks for is an **input** to the model rather than an output of it, and running the simulation teaches you the procedure but cannot measure the world.

Exercise	Simulation	What it gives, and what it does not
B.1	Oscilloscope	The whole exercise. The generator is silent, so step 4's *what changes in what you hear* needs a tone from Wave Composer or a real generator alongside
B.2	Standing Waves, *Standing Waves* screen	Rehearsal only. The speed of sound is what the model is given, not what it finds; the measurement needs a real tube and a real thermometer
B.3	Wave on a String	The harmonic sweep of step 2 and the shape of every mode. Tension is a slider without units, so the f_1 against $\sqrt{T}$ plot and the value of μ need the real apparatus
B.4	Resonance, *Single Oscillator* screen	The whole exercise, and better than the bench version: you set the mass, stiffness, and damping, so the Q you measure from the bandwidth can be checked against the Q those numbers predict
B.5	Wave Composer, *Analyzer* screen	Spectra and spectrograms of anything you play into the microphone, which is the body of the exercise. The listening test needs an editor to cut the attacks off
B.6	—	None. Levels, distances, and rooms are physical
B.7	Wave Composer, *Composer* screen	Steps 1 and 2: it sums up to four sinusoids and beat pairs and plays them, so the beats are both visible and countable. Steps 3 and 4 measure *you*, and need a partner
B.8	Wave Composer, *Analyzer* screen	The frequency measurement, by pitch tracking over roughly 60 to 800 Hz, which covers the middle octave step 2 asks for. The instrument itself has to be real
B.9	Standing Waves, *Instruments* screen	The prediction, exactly: the $1:2:3$ and $1:3:5$ ratios and the stopped pipe's octave drop. It deliberately omits the end correction, which is the quantity

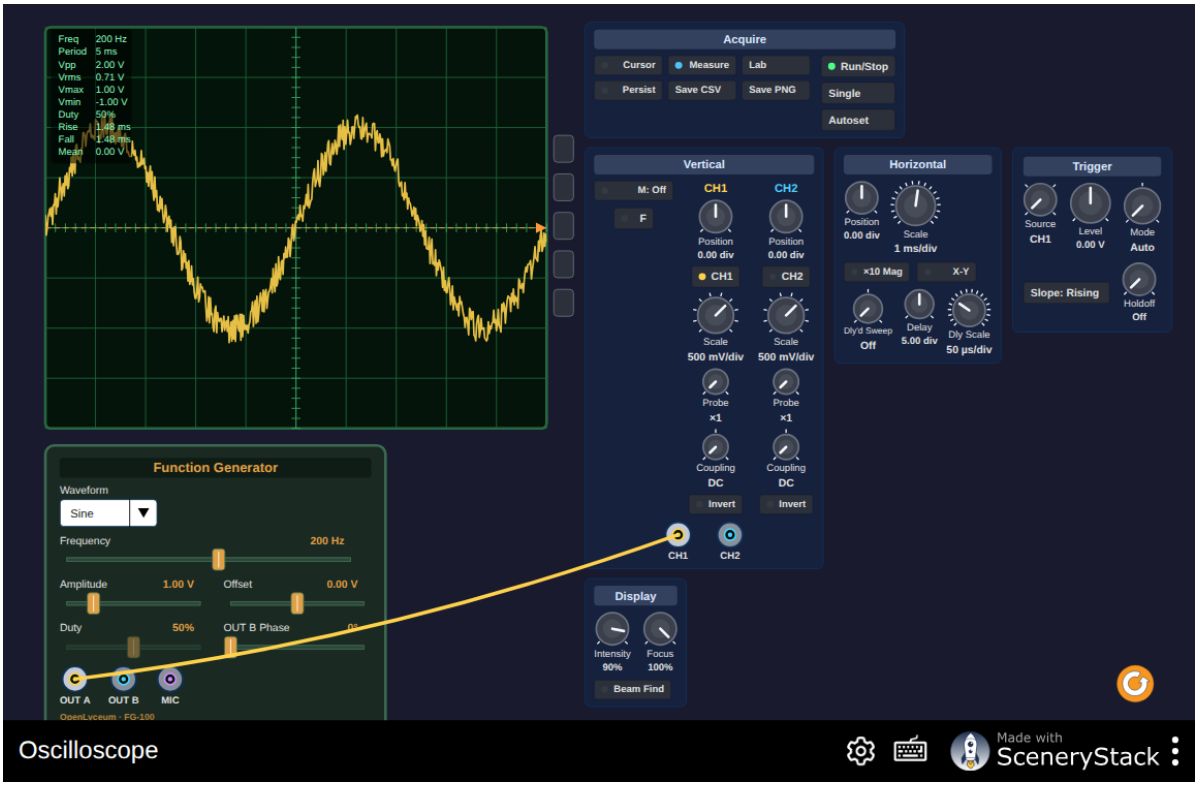


Figure 17.1: The bench of B.1, and of most of what follows. Its *Lab* menu holds four short drills, measuring V_{pp}, catching a normal trigger, finding the third harmonic in `FFT` mode, and matching a 90° Lissajous figure, which between them cover every control the exercises use. Work through those before touching a real instrument, and the real instrument stops being intimidating. Interactive simulation: Oscilloscope

> **Warning**
> Several of these exercises involve sustained tones through headphones. Set the level **before** you put them on, keep it at a comfortable conversational loudness, and do not use a tone to test "how loud you can stand". Chapter 6 explains why hearing damage is cumulative and permanent; this is a course where the equipment includes your ears.

17.2 How to Write These Up

Each exercise ends with an analysis section. A complete write-up has four parts, and the third is the one most often skipped:

1. **What you measured**: the raw numbers, with units, in a table.
2. **What you calculated**: the derived quantity, with the formula used.
3. **How well it agrees**: the percentage difference from the prediction, *and an estimate of your own uncertainty*. "Within 3%" means nothing until you say whether your measurement was good to 1% or to 10%.
4. **What the discrepancy means**, either the prediction is confirmed within uncertainty, or something is unaccounted for. Say which, and if the latter, name a candidate.

17.3 B.1 The Oscilloscope and the Function Generator

Supports Chapters 1–2. **Time** 60–90 minutes. **Simulation** Oscilloscope.

The point of the first exercise is not to discover anything. It is to learn to read the two instruments the rest of the exercises depend on, and to connect what you see on a screen to what you hear.

Procedure

1. Set the function generator to a 440 Hz sine wave and display it on the oscilloscope. Adjust the time base until two or three full cycles fill the screen.
2. Measure the period directly from the screen. Calculate the frequency and compare with the generator's setting.
3. Double the frequency to 880 Hz. Predict what happens to the trace before you press the button, then check.
4. Return to 440 Hz and halve the amplitude. Record what changes on the screen and what changes in what you hear.
5. Switch the generator to square, triangle, and sawtooth at the same 440 Hz. Sketch each waveform and describe in one sentence how each sounds compared with the sine.
6. Sing or play a sustained note into a microphone and display it. Identify the period on the screen and calculate the frequency.

Analysis

- From your measured period in step 2, what is your frequency and its uncertainty? Which is the limiting factor: the resolution of the time base, or your ability to judge where a cycle begins?
- Rank the four waveforms of step 5 from dullest to brightest. Chapter 5 will explain the ranking; record the order now, before you know the answer.
- The trace in step 6 is not a clean sinusoid. List two things visible in it that are not visible in the generator's sine.

17.4 B.2 The Speed of Sound

Supports Chapter 2. **Time** 60 minutes. **Simulation** Standing Waves, for rehearsal only.

Method A, Resonance in a tube

1. Hold a vibrating tuning fork of known frequency f over the open end of a tube whose length can be varied (a graduated cylinder being filled with water works well).
2. Find the shortest length L_1 at which the tube resonates loudly, and then the next length L_2.
3. The distance $L_2 - L_1$ is half a wavelength. Calculate λ and then $v = f\lambda$.

Using the *difference* of two resonance lengths rather than the first length alone eliminates the end correction, which is exactly what you want here and exactly what you will measure in B.9.

Method B, Time of flight

1. Place two microphones several meters apart, as far apart as your interface allows, and record both to the same file.
2. Make a sharp impulsive sound (a hand clap, two blocks struck together) beyond one microphone, on the line joining them.
3. Measure the delay between the two arrivals in the recording, and divide the separation by it.

Analysis

- Record the air temperature. Calculate the predicted speed of sound and compare.
- Which method gave the smaller uncertainty, and why?
- Method B is sensitive to where you stand. Explain why standing off the line joining the microphones makes the measured speed come out too high.

17.5 B.3 Standing Waves on a String

Supports Chapter 3. **Time** 90 minutes. **Simulation** Wave on a String, in part.

Procedure

1. Stretch a string horizontally from a mechanical vibrator over a pulley to a hanging mass, so that the tension is $T = mg$ and the vibrating length L is between the vibrator and the pulley.
2. Drive the string and sweep the frequency slowly upward from a few hertz. Record every frequency at which a clean standing-wave pattern appears, and the number of loops in each.
3. Repeat for at least four different hanging masses.
4. If a second string of different thickness is available, repeat step 2 for it at one fixed tension.

Analysis

- For one tension, plot f_n against n. The points should lie on a straight line through the origin. Do they? What is the slope, and what does it mean?
- Plot f_1 against $\sqrt{T}$ for your four masses. Is it linear? Extract the linear mass density μ from the slope and compare with the value you get by weighing a measured length of the string.
- The frequency at which a pattern looks *best* is not sharply defined. Estimate the width of the frequency range over which each resonance is visible, and use it to put an uncertainty on f_n.

17.6 B.4 Resonance and the Quality Factor

Supports Chapter 4. **Time** 60 minutes. **Simulation** Resonance.

Procedure

1. Choose a resonator: a wine glass, a length of pipe, a guitar body, or the string apparatus of B.3 held at one mode.
2. Drive it with a loudspeaker fed from the function generator, and measure the response amplitude with a microphone as you step the frequency through the resonance in small increments.
3. Take at least fifteen points, closely spaced near the peak.
4. Separately, excite the resonator impulsively (tap it) and record the decay.

Analysis

- Plot amplitude against frequency. Find the peak frequency f_0, and the two frequencies either side at which the amplitude has fallen to $1/\sqrt{2}$ of the peak. The difference between them is the bandwidth Δf; calculate $Q = f_0/\Delta f$.
- From your impulse recording, measure the time for the amplitude to fall to $1/e$ of its initial value. Estimate Q from the decay and compare with the value from the response curve.
- A wine glass has a Q in the hundreds; a guitar body's main resonance has a Q of order ten. Explain, in terms of the instrument's job, why a guitar should not have a high-Q body.

17.7 B.5 The Spectra of Musical Instruments

Supports Chapter 5. **Time** 90 minutes. **Simulation** Wave Composer.

This is the central exercise of the course.

Procedure

1. Record the **same pitch**, A3 at 220 Hz is convenient, played by at least four sources: a sung vowel, a plucked string, a bowed or blown sustained tone, and a pure sine from the generator. Record several seconds of steady tone from each, at roughly equal loudness.
2. For each recording, select a window from the *middle* of the sustained portion, well clear of the attack, and compute its spectrum.
3. Record the frequency and relative amplitude of the first eight partials of each.
4. Now record just the **first half-second** of each note, including the attack, and make a spectrogram of it.

Analysis

- Tabulate the partial amplitudes, in dB relative to the strongest partial, for all four sources. Which instrument has the strongest high harmonics? Does the ranking match the brightness you hear?
- Are the measured partial frequencies exact integer multiples of the fundamental? Report the deviation of the eighth partial as a percentage. Which source is least harmonic, and why might that be?
- From the spectrograms: describe how each note's spectrum changes during the first tenth of a second. Which partials arrive first?
- **The listening test.** Using an audio editor, cut the attack off each recording and play only the steady portions to a listener who has not seen which is which. How well can they identify the instruments? Relate the result to §5.4.

17.8 B.6 Sound Level, Distance, and Addition

Supports Chapter 7. **Time** 60 minutes. **Simulation** none; this one is physical.

Use a sound level meter if one is available. A phone app will do, provided every conclusion is drawn from *differences* between readings.

Procedure

1. Outdoors, or in the largest and least reverberant space available, set up a loudspeaker playing steady pink noise.

2. Measure the level at 1, 2, 4, and 8 meters on the axis of the speaker.

3. Return to a fixed distance. Measure the level with one speaker, then with two identical speakers side by side playing the *same* noise file, then with two playing *different* noise files.

4. Measure the level of several ordinary sounds: a quiet room, conversation at a meter, a hand clap, traffic.

Analysis

- Plot level against $\log_{10}$(distance). The slope should be -20 dB per decade: that is, -6 dB per doubling. What slope did you get? If it is shallower, what does that tell you about the room?
- In step 3, two speakers playing *different* noise should add to about +3 dB. What did two playing the *same* file give, and why is it different?
- Using the equal-loudness contours of §7.4, estimate the loudness level in phons of your measured traffic noise, and say what you had to assume about its spectrum to do so.

17.9 B.7 Beats and the Just-Noticeable Difference

Supports Chapter 8. **Time** 45 minutes. **Simulation** Wave Composer, for steps 1 and 2.

Procedure

1. Sound two tones together: one fixed at 440 Hz, one variable. Sum them electrically if you can, or use two speakers.
2. Set the second tone to 441, 443, 446, 450, 460, and 480 Hz in turn. For each, record: the beat rate you count, and a one-word description of the sensation (*beating*, *rough*, *two separate tones*).
3. Now find your own frequency discrimination. Have a partner alternate between 440 Hz and a slightly different frequency, and narrow the difference until you can no longer reliably tell which is which.
4. Repeat step 3 at 220 Hz and at 3520 Hz.

Analysis

- Plot your counted beat rate against the frequency difference. The prediction is a straight line of slope 1. Up to what difference does it hold, and what happens beyond?
- At what frequency difference did beating give way to roughness, and roughness to two separate tones? Compare the second of these with the critical bandwidth at 440 Hz quoted in §7.5.
- Express your discrimination threshold from steps 3 and 4 as a *percentage* of the center frequency at each of the three frequencies. Is the percentage constant? What does that tell you?

17.10 B.8 Measuring a Tuning

Supports Chapter 9. **Time** 90 minutes. **Simulation** Wave Composer, as the frequency meter.

Procedure

1. Choose a fixed-pitch instrument: a piano, an electronic keyboard, or a fretted guitar.
2. Record every note of one chromatic octave in the middle of its range, and measure the fundamental frequency of each to the nearest 0.1 Hz.
3. Also record notes two or three octaves apart, near the extremes of the instrument's range.
4. If a second instrument of a different kind is available, repeat.

Analysis

- Convert each measured frequency to cents above the lowest note of your octave. Tabulate alongside the equal-tempered values (0, 100, 200, …).
- Plot the deviation in cents from equal temperament for each note. Is the instrument equally tempered within your measurement uncertainty?
- Compare your measured major third and perfect fifth with both the just ratios (5:4 and 3:2) and the equal-tempered values. Which is the instrument closer to?
- From step 3: is the instrument's octave exactly 1200 cents? If you measured a piano, it very likely is not. Explain the result using §10.6.

17.11 B.9 Air Columns and the End Correction

Supports Chapter 11. **Time** 60 minutes. **Simulation** Standing Waves, for the prediction.

Procedure

1. Take a tube open at both ends. Measure its length and inside diameter.
2. Excite it, by blowing across one end, or with a loudspeaker at one end, and measure the frequencies of its first three resonances.
3. Close one end and repeat.
4. Repeat the whole procedure for a tube of noticeably different diameter but similar length.

Analysis

- For the open tube, are the resonances in the ratio 1: 2: 3? For the stopped tube, are they in the ratio 1: 3: 5? Report the measured ratios.
- Using the speed of sound at your measured temperature, calculate the *effective* length of each tube from its fundamental. Subtract the physical length to get the end correction.
- Divide your end correction by the tube radius. Compare with the theoretical value of about $0.6r$ per open end quoted in §11.2. Does the second tube give a consistent answer?
- A stopped tube of the same length sounds roughly an octave lower. Verify this from your measurements, and state the discrepancy.

17.12 B.10 Reverberation Time

Supports Chapter 14. **Time** 60 minutes, plus access to two very different rooms. **Simulation** none; this one needs two real rooms.

Procedure

1. Choose two rooms of markedly different character: a stairwell and a carpeted office, say. Measure the dimensions of each and estimate its volume.
2. In each, make a loud impulsive sound (a balloon burst is the classic; two hardwood blocks work) and record it.
3. From the recording, plot the decay of level in dB against time.
4. Repeat the measurement at three different positions in each room.

Analysis

- Measure T_{60}, the time for a 60 dB decay, by fitting a straight line to the decay curve. In most rooms you will not get a clean 60 dB above the noise floor; fit over the first 20 or 30 dB and extrapolate, and say that you did.
- Use the Sabine equation with your measured T_{60} and volume to find each room's total absorption, and then its average absorption coefficient. Is the value plausible for the surfaces you can see?

- Filter the recordings into octave bands and measure T_{60} in each. Is the reverberation time the same at 125 Hz as at 4 kHz? What does the difference say about the room's furnishings?
- How much did your three positions disagree? Is the reverberant field as uniform as the Sabine model assumes?

Chapter 18

The Final Project

The final project asks you to do two things that the problem sets cannot: to measure something nobody handed you the answer to, and to explain what you found to somebody who does not know any physics.

The second is the harder one, and it is the one most likely to matter to you after the course ends.

18.1 What the Project Is

Pick a question about musical sound that you can actually answer with a measurement, make the measurement, and present the result in ten minutes to an audience of your classmates, treated, for the purposes of the talk, as intelligent people who have not taken this course.

That is the whole specification. The constraints hiding inside it are worth making explicit.

It must be a question, not a topic. "The physics of the violin" is a topic, and a talk on a topic is a summary of things other people found out. "Does bowing closer to the bridge really add high harmonics, and how much?" is a question. You can be wrong about a question, which is what makes answering one worth ten minutes of an audience's attention.

You must measure something yourself. Not necessarily something new, almost certainly not something new, but something you recorded, timed, counted, or weighed. A project whose evidence is entirely other people's published figures is a literature review, and it does not meet the outcome.

The measurement must be possible with what you have. A phone, a laptop, and free software are enough for most of the topics below. If your question needs an anechoic chamber, choose another question, or find the version of it that a carpeted bedroom can answer.

18.2 Choosing a Question

The reliable move is to take something you already know from experience and ask *how much*. Three examples of the transformation:

Experience	Question you can measure
"My guitar sounds brighter near the bridge."	How does the amplitude of the 5th harmonic vary with plucking position, and does the measured pattern match the prediction that a pluck at L/n suppresses the nth harmonic?
"The stairwell makes my voice sound huge."	What is the stairwell's reverberation time, and how does it compare with my bedroom's? Does the Sabine equation predict it from the dimensions and surfaces?
"Cheap earbuds have no bass."	Where does the response of three pairs of headphones roll off at the low end, and by how much do they differ at 60 Hz?

Thirty starting points

Instruments. Plucking position and harmonic content · Why a wound bass string works · The inharmonicity of a real piano string, measured · How much a guitar's body resonance actually adds · What a mute does to a trumpet's spectrum · The clarinet's missing even harmonics, verified · Where a recorder's register hole has to go · The end correction of a real pipe · How a drum's pitch changes with head tension · Why a struck bar sounds different from a plucked string

Tuning and perception. Is a real piano stretched, and by how much · How close is a barbershop quartet to just intonation · How accurately can you tune a fifth by beats · Your own frequency discrimination against frequency · The missing fundamental, demonstrated on yourself · How many cents out does an interval have to be before you notice · Whether the dissonance of an interval changes with timbre

Voice. Formants of your own vowels, measured · How the vocal tract changes between a spoken and a sung vowel · Whether you can find your own singer's formant · How high a soprano can go before the vowel becomes unidentifiable

Rooms and recording. Reverberation time of three rooms · Room modes in a small bedroom · What a polar pattern really looks like · How much a lossy codec actually changes a waveform · Where aliasing becomes audible · Whether a loudspeaker's frequency response matches its specification · How stereo localization depends on delay versus level · What a hall's early reflections do to clarity

18.3 The Schedule

When	What
Week 4	Choose a question. Write it in one sentence.
Week 7	**Proposal**: one page: the question, how you will measure it, what result you expect, and what would count as the measurement failing.
Week 11	**Progress check**: show one real measurement, however rough.
Week 14	Draft slides; practice the talk out loud, timed.
Week 15	**Presentation**: ten minutes plus questions.

The Week 7 proposal exists because the most common way this project goes wrong is a question that cannot be measured, discovered in Week 13. Say in the proposal what would count as failure; if you cannot name a result that would disappoint you, the question probably has no content.

18.4 Giving the Talk

Ten minutes is about 1,200 words. That is very little, and it is enough, if you spend it on the right things.

Spend the first minute on the question, not on background. The audience does not need the history of acoustics. They need to know what you wanted to find out and why anyone would care. Play the sound if you have one, thirty seconds of a guitar plucked in two places does more than a slide about harmonics.

Show your own data, large. One graph you made, with axes labeled in units a non-specialist can read, beats six graphs from papers. If a number is the point of the talk, put that number on the slide by itself.

Say what you found, including if it was not what you expected. A measurement that disagreed with the prediction is a better talk than one that agreed, as long as you say what you think went wrong. Agreement is only interesting if the audience believed disagreement was possible.

Translate every technical term or drop it. You may use one or two, if you define them in plain words the first time. “The spectrum: that is, a list of which pure tones are present and how strong each one is” costs you four seconds and keeps the room with you. An undefined “spectral centroid” loses it permanently.

Do not read the slides. If a slide is readable in three seconds, it is doing its job.

18.5 What a Good Talk Sounds Like

Three tests, applied by the audience rather than by you:

- Can a listener state your question afterwards, in their own words?
- Can they say what you did to answer it?
- Do they believe the answer, and could they say why?

A talk that is technically flawless and fails all three has not met the communication outcome. This is worth saying plainly because it is a genuinely different skill from the one the problem sets develop, and it is the one that transfers.

18.6 Rubric

	Weight	What is assessed
The question	15%	Specific, answerable, and worth asking. Framed as something that could have come out otherwise.
The measurement	30%	Your own data. Method described well enough to repeat. Uncertainty estimated, not ignored.
The physics	25%	The result is connected to the physics of the course, correctly. A prediction is compared with the measurement and the comparison is honest.
The explanation	25%	A non-specialist can follow it. Terms defined or avoided. Figures legible and labeled. Time respected.
Response to questions	5%	Answers what was asked. Says "I don't know" where that is the true answer.

Note the weighting: the explanation is worth as much as the physics. That is deliberate, and it is the point of the exercise.

18.7 A Note on Honesty

Report what you measured, including the run that did not work and the data point you cannot explain. If your result disagrees with the textbook, say so and offer a reason: a systematic error in your method is a perfectly respectable finding, and identifying one is real physics. Quietly dropping the inconvenient half of your data is the one thing in this project that cannot be recovered from.

Chapter 19

Glossary

Terms are defined here as this book uses them. Where a word is used differently elsewhere, *overtone* and *partial* are the notorious pair, the entry says so.

Abbreviations that are merely expansions (sound pressure level, fast Fourier transform, ADSR, and the rest) are not repeated here; they are expanded on hover throughout the text.

Acronyms

ADSR attack, decay, sustain, release. 106, 338

HRTF head-related transfer function. 307, 308

ISO International Organization for Standardization. 141