

CONTENTS

DUTTON

Published by Penguin Group (USA) Inc.

375 Hudson Street, New York, New York 10014, U.S.A.

Penguin Group (Canada), 90 Eglinton Avenue East, Suite 700, Toronto, Ontario M4P 2Y3, Canada (a division of Pearson Penguin Canada Inc.); Penguin Books Ltd, 80 Strand, London WC2R 0RL, England; Penguin Ireland, 25 St Stephen's Green, Dublin 2, Ireland (a division of Penguin Books Ltd); Penguin Group (Australia), 250 Camberwell Road, Camberwell, Victoria 3124, Australia (a division of Pearson Australia Group Pty Ltd); Penguin Books India Pvt Ltd, 11 Community Centre, Panchsheel Park, New Delhi - 110 017, India; Penguin Group (NZ), cnr Airborne and Rosedale Roads, Albany, Auckland 1310, New Zealand (a division of Pearson New Zealand Ltd); Penguin Books (South Africa) (Pty) Ltd, 24 Sturdee Avenue, Rosebank, Johannesburg 2196, South Africa

Penguin Books Ltd, Registered Offices: 80 Strand, London WC2R 0RL, England

Published by Dutton, a member of Penguin Group (USA) Inc.

First printing, August 2006

20 19 18 17 16

Copyright © 2006 by Daniel J. Levitin

All rights reserved



REGISTERED TRADEMARK—MARCA REGISTRADA

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Levitin, Daniel J.

This is your brain on music : the science of a human obsession / Daniel J. Levitin.

p. cm.

Includes bibliographical references and index.

ISBN 0-525-94969-0 (hardcover)

1. Music—Psychological aspects. I. Title.

ML3830.L38 2006

781'.11—dc22

2006009055

Printed in the United States of America

Set in ITC Century Book

Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of both the copyright owner and the above publisher of this book.

The scanning, uploading, and distribution of this book via the Internet or via any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions, and do not participate in or encourage electronic piracy of copyrighted materials. Your support of the author's rights is appreciated.

While the author has made every effort to provide accurate telephone numbers and Internet addresses at the time of publication, neither the publisher nor the author assumes any responsibility for errors, or for changes that occur after publication. Further, the publisher does not have any control over and does not assume any responsibility for author or third-party Web sites or their content.

Introduction

I Love Music and I Love Science—Why Would I Want to Mix the Two?

1

1. What Is Music?

From Pitch to Timbre

13

2. Foot Tapping

Discerning Rhythm, Loudness, and Harmony

55

3. Behind the Curtain

Music and the Mind Machine

81

4. Anticipation

What We Expect From Liszt (and Ludacris)

109

5. You Know My Name, Look Up the Number

How We Categorize Music

129

6. After Dessert, Crick Was Still Four Seats Away from Me

Music, Emotion, and the Reptilian Brain

165

2. Foot Tapping

Discerning Rhythm, Loudness, and Harmony

I saw Sonny Rollins perform in Berkeley in 1977; he is one of the most melodic saxophone players of our time. Yet nearly thirty years later, while I can't remember any of the pitches that he played, I clearly remember some of the rhythms. At one point, Rollins improvised for three and a half minutes by playing the same one note over and over again with different rhythms and subtle changes in timing. All that power in one note! It wasn't his melodic innovation that got the crowd to their feet—it was rhythm. Virtually every culture and civilization considers movement to be an integral part of music making and listening. Rhythm is what we dance to, sway our bodies to, and tap our feet to. In so many jazz performances, the part that excites the audience most is the drum solo. It is no coincidence that making music requires the coordinated, rhythmic use of our bodies, and that energy be transmitted from body movements to a musical instrument. At a neural level, playing an instrument requires the orchestration of regions in our primitive, reptilian brain—the cerebellum and the brain stem—as well as higher cognitive systems such as the motor cortex (in the parietal lobe) and the planning regions of our frontal lobes, the most advanced region of the brain.

Rhythm, meter, and tempo are related concepts that are often confused with one another. Briefly, *rhythm* refers to the lengths of notes,

tempo refers to the pace of a piece of music (the rate at which you would tap your foot to it), and *meter* refers to when you tap your foot hard versus light, and how these hard and light taps group together to form larger units.

One of the things we usually want to know when performing music is how long a note is to be played. The relationship between the length of one note and another is what we call rhythm, and it is a crucial part of what turns sounds into music. Among the most famous rhythms in our culture is the rhythm often called “shave-and-a-haircut, two bits,” sometimes used as the “secret” knock on a door. An 1899 recording by Charles Hale, “At a Darktown Cakewalk,” is the first documented use of this rhythm. Lyrics were later attached to the rhythm in a song by Jimmie Monaco and Joe McCarthy called “Bum-Diddle-De-Um-Bum, That’s It!” in 1914. In 1939, the same musical phrase was used in the song “Shave and a Haircut—Shampoo” by Dan Shapiro, Lester Lee, and Milton Berle. How the word *shampoo* became *two-bits* is a mystery. Even Leonard Bernstein got into the act by scoring this rhythm in the song “Gee, Officer Krupke” from the musical *West Side Story*. In “shave-and-a-haircut” we hear a series of notes of two different lengths, long and short; the long notes are twice as long as the short ones: long-short-short-long-long (rest) long-long.

In the *William Tell* overture by Rossini (what many of us know as the theme from *The Lone Ranger*) we also hear a series of notes of two different lengths, long and short; again, the long notes are twice as long as the short ones: da-da-bump da-da-bump da-da-bump bump bump (here I’ve used the “da” syllable for short, and the “bump” syllable for long). “Mary Had a Little Lamb” uses short and long syllables, too, in this case six equal duration notes (Ma-ry had a lit-tle) followed by a long one (lamb) roughly twice as long as the short ones. The rhythmic ratio of 2:1, like the octave in pitch ratios, appears to be a musical universal. We see it in the theme from *The Mickey Mouse Club* (bump-ba bump-ba bump-ba bump-ba bump-ba bump-ba baaaaah) in which we have three levels of duration, each one twice as long as the other. We see it in The Police’s “Every Breath You Take” (da-da-bump da-da baaaaah), in which there are again three levels:

Ev-ry breath you-oo taaake

1 1 2 2 4

(The 1 represents one unit of some arbitrary time just to illustrate that the words *breath* and *you* are twice as long as the syllables *Ev* and *ry*, and that the word *take* is four times as long as *Ev* or *ry* and twice as long as *breath* or *you*.)

Rhythms in most of the music we listen to are seldom so simple. In the same way that a particular arrangement of pitches—the scale—can evoke music of a different culture, style, or idiom, so can a particular arrangement of rhythms. Although most of us couldn’t reproduce a complex Latin rhythm, we recognize as soon as we hear it that it is Latin, as opposed to Chinese, Arabic, Indian, or Russian. When we organize rhythms into strings of notes, of varying lengths and emphases, we develop meter and establish tempo.

Tempo refers to the pace of a musical piece—how quickly or slowly it goes by. If you tap your foot or snap your fingers in time to a piece of music, the tempo of the piece will be directly related to how fast or slow you are tapping. If a song is a living, breathing entity, you might think of the tempo as its gait—the rate at which it walks by—or its pulse—the rate at which the heart of the song is beating. The word *beat* indicates the basic unit of measurement in a musical piece; this is also called the *tactus*. Most often, this is the natural point at which you would tap your foot or clap your hands or snap your fingers. Sometimes, people tap at half or twice the beat, due to different neural processing mechanisms from one person to another as well as differences in musical background, experience, and interpretation of a piece. Even trained musicians can disagree on what the tapping rate should be. But they always agree on the underlying speed at which the piece is unfolding, also called tempo; the disagreements are simply about subdivisions or superdivisions of that underlying pace.

Paula Abdul’s “Straight Up” and AC/DC’s “Back in Black” have a tempo of 96, meaning that there are 96 beats per minute. If you dance to “Straight Up” or “Back in Black,” it is likely that you will be putting a foot

down 96 times per minute or perhaps 48, but not 58 or 69. In “Back in Black” you can hear the drummer playing a beat on his high-hat cymbal at the very beginning, steadily, deliberately, at precisely 96 beats per minute. Aerosmith’s “Walk This Way” has a tempo of 112, Michael Jackson’s “Billie Jean” has a tempo of 116, and the Eagles’ “Hotel California” has a tempo of 75.

Two songs can have the same tempo but feel very different. In “Back in Black,” the drummer plays his cymbal twice for every beat (eighth notes) and the bass player plays a simple, syncopated rhythm perfectly in time with the guitar. On “Straight Up” there is so much going on, it is difficult to describe it in words. The drums play a complex, irregular pattern with beats as fast as sixteenth notes, but not continuously—the “air” between drum hits imparts a sound typical of funk and hip-hop music. The bass plays a similarly complex and syncopated melodic line that sometimes coincides with and sometimes fills in the holes of the drum part. In the right speaker (or the right ear of headphones) we hear the only instrument that actually plays on the beat every beat—a Latin instrument called an *afuche* or *cabasa* that sounds like sandpaper or beans shaking inside a gourd. Putting the most important rhythm on a light, high-pitched instrument is an innovative rhythmic technique that turns upside down the normal rhythmic conventions. While all this is going on, synthesizers, guitar, and special percussion effects fly in and out of the song dramatically, emphasizing certain beats now and again to add excitement. Because it is hard to predict or memorize where many of these are, the song holds a certain appeal over many, many listenings.

Tempo is a major factor in conveying emotion. Songs with fast tempos tend to be regarded as happy, and songs with slow tempos as sad. Although this is an oversimplification, it holds in a remarkable range of circumstances, across many cultures, and across the lifespan of an individual. The average person seems to have a remarkable memory for tempo. In an experiment that Perry Cook and I published in 1996, we asked people to simply sing their favorite rock and popular songs from memory and we were interested to know how close they came to the actual tempo of the recorded versions of those songs. As a baseline, we

considered how much variation in tempo the average person can detect; that turns out to be 4 percent. In other words, for a song with a tempo of 100 bpm, if the tempo varies between 96–100, most people, even some professional musicians, won’t detect this small change (although most drummers would—their job requires that they be more sensitive to tempo than other musicians, because they are responsible for maintaining tempo when there is no conductor to do it for them). A majority of people in our study—nonmusicians—were able to sing songs within 4 percent of their nominal tempo.

The neural basis for this striking accuracy is probably in the cerebellum, which is believed to contain a system of timekeepers for our daily lives and to synchronize to the music we are hearing. This means that somehow, the cerebellum is able to remember the “settings” it uses for synchronizing to music as we hear it, and it can recall those settings when we want to sing a song from memory. It allows us to synchronize our singing with a memory of the last time we sang. The basal ganglia—what Gerald Edelman has called “the organs of succession”—are almost certainly involved, as well, in generating and shaping rhythm, tempo, and meter.

Meter refers to the way in which the pulses or beats are grouped together. Generally when we’re tapping or clapping along with music, there are some beats that we feel more strongly than others. It feels as if the musicians play this beat louder and more heavily than the others. This louder, heavier beat is perceptually dominant, and other beats that follow it are perceptually weaker until another strong one comes in. Every musical system that we know of has patterns of strong and weak beats. The most common pattern in Western music is for the strong beats to occur once every 4 beats: STRONG-weak-weak-weak STRONG-weak-weak-weak. Usually the third beat in a four-beat pattern is somewhat stronger than the second and fourth: There is a hierarchy of beat strengths, with the first being the strongest, the third being next, followed by the second and fourth. Somewhat less often the strong beat occurs once in every three in what we call the “waltz” beat: STRONG-weak-weak STRONG-weak-weak. We usually count to these beats as

well, in a way that emphasizes which one is the strong beat: ONE-two-three-four, ONE-two-three-four, or ONE-two-three, ONE-two-three.

Of course music would be boring if we only had these straight beats. We might leave one out to add tension. Think of “Twinkle, Twinkle Little Star,” written by Mozart when he was six years old. The notes don’t occur on every beat:

ONE-two-three-four
ONE-two-three-(rest)
ONE-two-three-four
ONE-two-three-(rest):

TWIN-kle twin-kle
LIT-tle star (rest)
HOW-I won-der
WHAT you are (rest).

A nursery rhyme written to this same tune, “Ba Ba Black Sheep” subdivides the beat. A simple ONE-two-three-four can be divided into smaller, more interesting parts:

BA ba black sheep
HAVE-you-any-wool?

Notice that each syllable in “have-you-any” goes by twice as fast as the syllables “ba ba black.” The quarter notes have been divided in half, and we can count this as

ONE-two-three-four
ONE-and-two-and-three-(rest).

In “Jailhouse Rock,” performed by Elvis Presley and written by two outstanding songwriters of the rock era, Jerry Leiber and Mike Stoller,

the strong beat occurs on the first note Presley sings, and then every fourth note after that:

[Line 1:] WAR-den threw a party at the
[Line 2:] COUN-ty jail (rest) the
[Line 3:] PRIS-on band was there and they be-
[Line 4:] GAN to wail

In music with lyrics, the words don’t always line up perfectly with the downbeats; in “Jailhouse Rock” part of the word *began* starts before a strong beat and finishes on that strong beat. Most nursery rhymes and simple folk songs, such as “Ba Ba Black Sheep” or “Frère Jacques,” don’t do this. This lyrical technique works especially well on “Jailhouse Rock” because in speech the accent is on the second syllable of *began*; spreading the word across lines like this gives the song additional momentum.

By convention in Western music, we have names for the note durations similar to the way we name musical intervals. A musical interval of a “perfect fifth” is a relative concept—it can start on any note, and then by definition, notes that are either seven semitones higher or seven semitones lower in pitch are considered a perfect fifth away from the starting note. The standard duration is called a whole note and it lasts four beats, regardless of how slow or how fast the music is moving—that is, irrespective of tempo. (At a tempo of sixty beats per minute—as in the *Funeral March*—each beat lasts one second, so a whole note would last four seconds.) A note with half the duration of a whole note is called, logically enough, a half note, and a note half as long as that is called a quarter note. For most music in the popular and folk tradition, the quarter note is the basic pulse—the four beats that I was referring to earlier are beats of a quarter note. We talk about such songs as being in 4/4 time: The numerator tells us that the song is organized into groups of four notes, and the denominator tells us that the basic note length is a quarter

note. In notation and conversation, we refer to each of these groups of four notes as a measure or a bar. One measure of music in 4/4 time has four beats, where each beat is a quarter note. This does not imply that the only note duration in the measure is the quarter note. We can have notes of any duration, or rests—that is to say, no notes at all; the 4/4 indication is only meant to describe how we count the beats.

“Ba Ba Black Sheep” has four quarter notes in its first measure, and then eighth notes (half the duration of a quarter note) and a quarter note rest in the second measure. I’ve used the symbol | to indicate a quarter note, and L to indicate an eighth note, and I’ve kept the spacing between syllables proportional to how much time is spent on them:

[measure 1:] ba ba black sheep
 | | | |
 [measure 2:] have you an- y wool (rest)
 L L L L | |

You can see in the diagram that the eighth notes go by twice as fast as the quarter notes.

In “That’ll Be the Day” by Buddy Holly, the song begins with a pickup note; the strong beat occurs on the next note and then every fourth note after that, just as in “Jailhouse Rock”:

THAT’ll be the day (rest) when
YOU say good-bye-yes;
THAT’ll be the day (rest) when
YOU make me cry-hi; you
SAY you gonna leave (rest) you
KNOW it’s a lie ’cause
THAT’ll be the day-ay-
AY when I die.

Notice how, like Elvis, Holly cuts a word in two across lines (*day* in the last two lines). To most people, the tactus is four beats between

downbeats of this song, and they would tap their feet four times from one downbeat to the next. Here, all caps indicate the downbeat as before, and **bold** indicates when you would tap your foot against the floor:

Well

THAT’ll be the day (rest) when

YOU say good-bye-yes;

THAT’ll be the day (rest) when

YOU make me cry-hi; you

SAY you gonna leave (rest) you

KNOW it’s a lie ’cause

THAT’ll be the day-ay-

AY when I die.

If you pay close attention to the song’s lyrics and their relationship to the beat, you’ll notice that a foot tap occurs in the middle of some of the beats. The first *say* on the second line actually begins before you put your foot down—your foot is probably up in the air when the word *say* starts, and you put your foot down in the middle of the word. The same thing happens with the word *yes* later in that line. Whenever a note anticipates a beat—that is, when a musician plays a note a bit earlier than the strict beat would call for—this is called syncopation. This is a very important concept that relates to expectation, and ultimately to the emotional impact of a song. The syncopation catches us by surprise, and adds excitement.

As with many songs, some people feel “That’ll Be the Day” in half time; there’s nothing wrong with this—it is another interpretation and a valid one—and they tap their feet twice in the same amount of time other people tap four times: once on the downbeat, and again two beats later.

The song actually begins with the word *Well* that occurs before a strong beat—this is called a pickup note. Holly uses two words, *Well*, *you*, as pickup notes to the verse, also, and then right after them we’re in sync again with the downbeats:

[pick up] Well, you
 [line 1] GAVE me all your lovin' and your
 [line 2] (REST) tur-tle dovin' (rest)
 [line 3] ALL your hugs and kisses and your
 [line 4] (REST) money too.

What Holly does here that is so clever is that he violates our expectations not just with anticipations, but by delaying words. Normally, there would be a word on every downbeat, as in children's nursery rhymes. But in lines two and four of the song, the downbeat comes and he's silent! This is another way that composers build excitement, by not giving us what we would normally expect.

When people clap their hands or snap their fingers with music, they sometimes quite naturally, and without training, keep time differently than they would do with their feet: They clap or snap not on the downbeat, but on the second beat and the fourth beat. This is the so-called backbeat that Chuck Berry sings about in his song "Rock and Roll Music."

John Lennon said that the essence of rock and roll songwriting for him was to "Just say what it is, simple English, make it rhyme, and put a backbeat on it." In "Rock and Roll Music" (which John sang with the Beatles), as on most rock songs, the backbeat is what the snare drum is playing: The snare drum plays only on the second and fourth beat of each measure, in opposition to the strong beat which is on one, and a secondary strong beat, on three. This backbeat is the typical rhythmic element of rock music, and Lennon used it a lot as in "Instant Karma" (*whack* below indicates where the snare drum is played in the song, on the backbeat):

Instant karma's gonna get you
 (rest) *whack* (rest) *whack*
 "Gonna knock you right on the head"
 (rest) *whack* (rest) *whack*

...

*But we all *whack* shine *whack**
*on *whack* (rest) *whack**
*Like the moon *whack* and the stars *whack**
*and the sun *whack* (rest) *whack**

In "We Will Rock You" by Queen, we hear what sounds like feet stamping on stadium bleachers twice in a row (boom-boom) and then hand-clapping (CLAP) in a repeating rhythm: boom-boom-CLAP, boom-boom-CLAP; the CLAP is the backbeat.

Imagine now the John Philip Sousa march, "The Stars and Stripes Forever." If you can hear it in your mind, you can tap your foot along with the mental rhythm. While the music goes "DAH-dah-ta DUM-dum dah DUM-dum dum-dum DUM," your foot will be tapping DOWN-up DOWN-up DOWN-up DOWN-up. In this song, it is natural to tap your foot for every two quarter notes. We say that this song is "in two," meaning that the natural grouping of rhythms is two quarter notes per beat.

Now imagine "My Favorite Things" (words and music by Richard Rodgers and Oscar Hammerstein). This song is in waltz time, or what is called 3/4 time. The beats seem to arrange themselves in groups of three, with a strong beat followed by two weak ones. "RAIN-drops-on ROSE-es and WHISK-ers-on KIT-tens (rest)." ONE-two-three ONE-two-three ONE-two-three ONE-two-three.

As with pitch, small-integer ratios of durations are the most common, and there is accumulating evidence that they are easier to process neurally. But, as Eric Clarke notes, small-integer ratios are almost never found in samples of real music. This indicates that there is a quantization process—equalizing durations—occurring during our neural processing of musical time. Our brains treat durations that are similar as being equal, rounding some up and some down in order to treat them as simple integer ratios such as 2:1, 3:1 and 4:1. Some musics use more complex ratios than these; Chopin and Beethoven use nominal ratios of 7:4 and 5:4 in some of their piano works, in which seven or five notes are played with one hand while the other hand plays four. As with pitch, any ratio is

theoretically possible, but there are limitations to what we can perceive and remember, and there are limitations based on style and convention.

The three most common meters in Western music are: 4/4, 2/4, and 3/4. Other rhythmic groupings exist, such as 5/4, 7/4, and 9/4. A somewhat common meter is 6/8, in which we count six beats to a measure, and each eighth note gets one beat. This is similar to 3/4 waltz time, the difference being that the composer intends for the musicians to “feel” the music in groups of six rather than groups of three, and for the underlying pulse to be the shorter-duration eighth note rather than a quarter note. This points to the hierarchy that exists in musical groupings. It is possible to count 6/8 as two groups of 3/8 (ONE-two-three ONE-two-three) or as one group of six (ONE-two-three-FOUR-five-six) with a secondary accent on the fourth beat, and to most listeners these are uninteresting subtleties that only concern a performer. But there may be brain differences. We know that there are neural circuits specifically related to detecting and tracking musical meter, and we know that the cerebellum is involved in setting an internal clock or timer that can synchronize with events that are out-there-in-the-world. No one has yet done the experiment to see if 6/8 and 3/4 have different neural representations, but because musicians truly treat them as different, there is a high probability that the brain does also. A fundamental principle of cognitive neuroscience is that the brain provides the biological basis for any behaviors or thoughts that we experience, and so at some level there must be neural differentiation wherever there is behavioral differentiation.

Of course, 4/4 and 2/4 time are easy to walk to, dance to, or march to because (since they are even numbers) you always end up with the same foot hitting the floor on a strong beat. Three-quarter is less natural to walk to; you’ll never see a military outfit or infantry division marching to 3/4. Five-quarter time is used once in a while, the most famous examples being Lalo Shiffrin’s theme from *Mission: Impossible*, and the Dave Brubeck song “Take Five.” As you count the pulse and tap your foot to these songs, you’ll see that the basic rhythms group into fives: ONE-two-three-four-five, ONE-two-three-four-five. There is a secondary strong beat in Brubeck’s composition on the four: ONE-two-three-FOUR-five. In

this case, many musicians think of 5/4 beats as consisting of alternating 3/4 and 2/4 beats. In “Mission: Impossible,” there is no clear subdivision of the five. Tchaikovsky uses 5/4 time for the second movement of his Sixth Symphony. Pink Floyd used 7/4 for their song “Money,” as did Peter Gabriel for “Salisbury Hill”; if you try to tap your foot or count along, you’ll need to count seven between each strong beat.

I left discussion of loudness for almost-last, because there really isn’t much to say about loudness in terms of definition that most people don’t already know. One counterintuitive point is that loudness is, like pitch, an entirely psychological phenomenon, that is, loudness doesn’t exist in the world, it only exists in the mind. And this is true for the same reason that pitch only exists in the mind. When you’re adjusting the output of your stereo system, you’re technically increasing the amplitude of the vibration of molecules, which in turn is interpreted as loudness by our brains. The point here is that it takes a brain to experience what we call “loudness.” This may seem largely like a semantic distinction, but it is important to keep our terms straight. Several odd anomalies exist in the mental representation of amplitude, such as loudnesses not being additive the way that amplitudes are (loudness, like pitch, is logarithmic), or the phenomenon that the pitch of a sinusoidal tone varies as a function of its amplitude, or the finding that sounds can appear to be louder than they are when they have been electronically processed in certain ways—such as dynamic range compression—that are often done in heavy metal music.

Loudness is measured in decibels (named after Alexander Graham Bell and abbreviated dB) and it is a dimensionless unit like percent; it refers to a ratio of two sound levels. In this sense, it is similar to talking about musical intervals, but not to talking about note names. The scale is logarithmic, and doubling the intensity of a sound source results in a 3 dB increase in sound. The logarithmic scale is useful for discussing sound because of the ear’s extraordinary sensitivity: The ratio between the loudest sound we can hear without causing permanent damage and the softest sound we can detect is a million to one, when measured as

sound-pressure levels in the air; on the dB scale this is 120 dB. The range of loudnesses we can perceive is called the dynamic range. Sometimes critics talk about the dynamic range that is achieved on a high-quality music recording; if a record has a dynamic range of 90 dB, it means that the difference between the softest parts on the record and the loudest parts is 90 dB—considered high fidelity by most experts, and beyond the capability of most home audio systems.

Our ears compress sounds that are very loud in order to protect the delicate components of the middle and inner ear. Normally, as sounds get louder in the world, our perception of the loudness increases proportionately to them. But when sounds are really loud, a proportional increase in the signal transmitted by the eardrum would cause irreversible damage. The compression of the sound levels—of the dynamic range—means that large increases in sound level in the world create much smaller changes of level in our ears. The inner hair cells have a dynamic range of 50 decibels (dB) and yet we can hear over a 120 dB dynamic range. For every 4 dB increase in sound level, a 1 dB increase is transmitted to the inner hair cells. Most of us can detect when this compression is taking place; compressed sounds have a different quality.

Acousticians have developed a way to make it easy to talk about sound levels in the environment—because dBs express a ratio between two values, they chose a standard reference level (20 micropascals of sound pressure) which is approximately equal to the threshold of human hearing for most healthy people—the sound of a mosquito flying ten feet away. To avoid confusion, when decibels are being used to reflect this reference point of sound pressure level, we refer to them as dB (SPL). Here are some landmarks for sound levels, expressed in dB (SPL):

0 dB	Mosquito flying in a quiet room, ten feet away from your ears
20 dB	A recording studio or a very quiet executive office
35 dB	A typical quiet office with the door closed and computers off
50 dB	Typical conversation in a room

75 dB	Typical, comfortable music listening level in headphones
100–105 dB	Classical music or opera concert during loud passages; some portable music players go to 105 dB
110 dB	A jackhammer three feet away
120 dB	A jet engine heard on the runway from three hundred feet away; a typical rock concert
126–130 dB	Threshold of pain and damage; a rock concert by the Who (note that 126 dB is four times as loud as 120 dB)
180 dB	Space shuttle launch
250–275 dB	Center of a tornado; volcanic eruption

Conventional foam insert earplugs can block about 25 dB of sound, although they do not do so across the entire frequency range. Earplugs at a Who concert can minimize the risk of permanent damage by bringing down the levels that reach the ear close to 100–110 dB (SPL). The over-the-ear type of ear protector worn at rifle firing ranges and by airport landing personnel is often supplemented by in-the-ear plugs to afford maximum protection.

A lot of people like really loud music. Concertgoers talk about a special state of consciousness, a sense of thrills and excitement, when the music is really loud—over 115 dB. We don't yet know why this is so. Part of the reason may be related to the fact that loud music saturates the auditory system, causing neurons to fire at their maximum rate. When many, many neurons are maximally firing, this could cause an emergent property, a brain state qualitatively different from when they are firing at normal rates. Still, some people like loud music, and some people don't.

Loudness is one of the seven major elements of music along with pitch, rhythm, melody, harmony, tempo, and meter. Very tiny changes in loudness have a profound effect on the emotional communication of music. A pianist may play five notes at once and make one note only slightly louder than the others, causing it to take on an entirely different role in our overall perception of the musical passage. Loudness is also an important cue to rhythms, as we saw above, and to meter, because it is the loudness of notes that determines how they group rhythmically.

* * *

Now we have come full circle and return to the broad subject of pitch. Rhythm is a game of expectation. When we tap our feet we are predicting what is going to happen in the music next. We also play a game of expectations in music with pitch. Its rules are key and harmony. A musical key is the tonal context for a piece of music. Not all musics have a key. African drumming, for instance, doesn't, nor does the twelve-tone music of contemporary composers such as Schönberg. But virtually all of the music we listen to in Western culture—from commercial jingles on the radio to the most serious symphony by Bruckner, from the gospel music of Mahalia Jackson to the punk of the Sex Pistols—has a central set of pitches that it comes back to, a tonal center, the key. The key can change during the course of the song (called modulation), but by definition, the key is generally something that holds for a relatively long period of time during the course of the song, typically on the order of minutes.

If a melody is based on the C major scale, for example, we generally say that the melody is “in the key of C.” This means that the melody has a momentum to return to the note C, and that even if it doesn't end on a C, the note C is what listeners are keeping in their minds as the dominant and focal note of the entire piece. The composer may temporarily use notes from outside the C major scale, but we recognize those as departures—something like a quick edit in a movie to a parallel scene or a flashback, in which we know that a return to the main plotline is imminent and inevitable. (For a more detailed look at music theory see Appendix 2.)

The attribute of pitch in music functions within a scale or a tonal/harmonic context. A note doesn't always sound the same to us every time we hear it: We hear it within the context of a melody and what has come before, and we hear it within the context of the harmony and chords that are accompanying it. We can think of it like flavor: Oregano tastes good with eggplant or tomato sauce, maybe less good with banana pudding. Cream takes on a different gustatory meaning when it is on top of strawberries from when it is in coffee or part of a creamy garlic salad dressing.

In “For No One” by the Beatles, the melody is sung on one note for

two measures, but the chords accompanying that note change, giving it a different mood and a different sound. The song “One Note Samba” by Antonio Carlos Jobim actually contains many notes, but one note is featured throughout the song with changing chords accompanying it, and we hear a variety of different shades of musical meaning as this unfolds. In some chordal contexts, the note sounds bright and happy, in others, pensive. Another thing that most of us are expert in, even if we are non-musicians, is recognizing familiar chord progressions, even in the absence of the well-known melody. Whenever the Eagles play this chord sequence in concert

B minor / F-sharp major / A major / E major / G major / D major /
E minor / F-sharp major

they don't have to play more than three chords before thousands of non-musician fans in the audience know that they are going to play “Hotel California.” And even as they have changed the instrumentation over the years, from electric to acoustic guitars, from twelve-string to six-string guitars, people recognize those chords; we even recognize them when they're played by an orchestra coming out of cheap speakers in a Muzak version in the dentist's office.

Related to the topic of scales and major and minor is the topic of tonal consonance and dissonance. Some sounds strike us as unpleasant, although we don't always know why. Fingernails screeching on a chalkboard are a classic example, but this seems to be true only for humans; monkeys don't seem to mind (or at least in the one experiment that was done, they like that sound as much as they like rock music). In music, some people can't stand the sound of distorted electric guitars; others won't listen to anything else. At the harmonic level—that is, the level of the notes, rather than the timbres involved—some people find particular intervals or chords particularly unpleasant. Musicians refer to the pleasing-sounding chords and intervals as consonant and the unpleasing ones as dissonant. A great deal of research has focused on the problem of why we find consonant some intervals and not others, and there is

currently no agreement about this. So far, we've been able to figure out that the brain stem and the dorsal cochlear nucleus—structures that are so primitive that all vertebrates have them—can distinguish between consonance and dissonance; this distinction happens before the higher level, human brain region—the cortex—gets involved.

Although the neural mechanisms underlying consonance and dissonance are debated, there is widespread agreement about some of the intervals that are deemed consonant. A unison interval—the same note played with itself—is deemed consonant, as is an octave. These create simple integer frequencies ratios of 1:1 and 2:1 respectively. (From an acoustics standpoint, half of the peaks in the waveform for octaves line up with each other perfectly, the other half fall exactly in between two peaks.) Interestingly, if we divide the octave precisely in half, the interval we end up with is called a tritone and most people find it the most disagreeable interval possible. Part of the reason for this may be related to the fact that the tritone does not come from a simple integer ratio, its ratio being 43:32. We can look at consonance from an integer ratio perspective. A ratio of 3:1 is a simple integer ratio, and that defines two octaves. A ratio of 3:2 is also a simple integer ratio, and that defines the interval of a perfect fifth. This is the distance between, for example, C and the G above it. The distance from that G to the C above it forms an interval of a perfect fourth, and its frequency ratio is 4:3.

The particular notes found in our major scale trace their roots back to the ancient Greeks and their notions of consonance. If we start with a note C and simply add the interval of a perfect fifth to it iteratively, we end up generating a set of frequencies that are very close to the current major scale: C - G - D - A - E - B - F-sharp - C-sharp - G-sharp - D-sharp - A-sharp - E-sharp (or F), and then back to C. This is known as the circle of fifths because after going through the cycle, we end up back at the note we started on. Interestingly, if we follow the overtone series, we can generate frequencies that are somewhat close to the major scale as well.

A single note cannot, by itself, be dissonant, but it can sound dissonant against the backdrop of certain chords, particularly when the chord implies a key that the single note is not part of. Two notes can sound dis-

sonant together, both when played simultaneously or in sequence, if the sequence does not conform to the customs we have learned that go with our musical idioms. Chords can also sound dissonant, especially when they are drawn from outside the key that has been established. Bringing all these factors together is the task of the composer. Most of us are very discriminating listeners, and when the composer gets the balance just slightly wrong, our expectations have been betrayed more than we can stand, and we switch radio stations, pull off the earphones, or just walk out of the room.

I've reviewed the major elements that go into music: pitch, timbre, key, harmony, loudness, rhythm, meter, and tempo. Neuroscientists deconstruct sound into its components to study selectively which brain regions are involved in processing each of them, and musicologists discuss their individual contributions to the overall aesthetic experience of listening. But music—real music—succeeds or fails because of the relationship among these elements. Composers and musicians rarely treat these in total isolation; they know that changing a rhythm may also require changing pitch or loudness, or the chords that accompany that rhythm. One approach to studying the relationship between these elements traces its origins back to the late 1800s and the Gestalt psychologists.

In 1890, Christian von Ehrenfels was puzzled by something all of us take for granted and know how to do: melodic transposition. Transposition is simply singing or playing a song in a different key or with different pitches. When we sing "Happy Birthday" we just follow along with the first person who started singing, and in most cases, this person just starts on any note that she feels like. She might even have started on a pitch that is not a recognized note of the musical scale, falling between, say, C and C-sharp, and almost no one would notice or care. Sing "Happy Birthday" three times in a week and you might be singing three completely different sets of pitches. Each version of the song is called a transposition of the others.

The Gestalt psychologists—von Ehrenfels, Max Wertheimer, Wolfgang Köhler, Kurt Koffka, and others—were interested in the problem of

configurations, that is, how it is that elements come together to form wholes, objects that are qualitatively different from the sum of their parts, and cannot be understood in terms of their parts. The word *Gestalt* has entered the English language to mean a unified whole form, applicable to both artistic and nonartistic objects. One can think of a suspension bridge as a Gestalt. The functions and utility of the bridge are not easily understood by looking at pieces of cable, girders, bolts, and steel beams; it is only when they come together in the form of a bridge that we can apprehend how a bridge is different from, say, a construction crane that might be made out of the same parts. Similarly, in painting, the relationship between elements is a critical aspect of the final artistic product. The classic example is a face—the *Mona Lisa* would not be what it is if the eyes, nose, and mouth were painted entirely as they are but were scattered across the canvas in a different arrangement.

The Gestaltists wondered how it is that a melody—composed of a set of specific pitches—could retain its identity, its recognizability, even when all of its pitches are changed. Here was a case for which they could not generate a satisfying theoretical explanation, the ultimate triumph of form over detail, of the whole over the parts. Play a melody using any set of pitches, and so long as the relation between those pitches is held constant, it is the same melody. Play it on different instruments and people still recognize it. Play it at half speed or double speed, or impose all of these transformations at the same time, and people still have no trouble recognizing it as the original song. The influential Gestalt school was formed to address this particular question. Although they never answered it, they did go on to contribute enormously to our understanding of how objects in the visual world are organized, through a set of rules that are taught in every introductory psychology class, the “Gestalt Principles of Grouping.”

Albert Bregman, a cognitive psychologist at McGill University, has performed a number of experiments over the last thirty years to develop a similar understanding of grouping principles for sound. The music theorist Fred Lerdahl from Columbia University and the linguist Ray Jack-

endoff from Brandeis University (now at Tufts University) tackled the problem of describing a set of rules, similar to the rules of grammar in spoken language, that govern musical composition, and these include grouping principles for music. The neural basis for these principles has not been completely worked out, but through a series of clever behavioral experiments we have learned a great deal about the phenomenology of the principles.

In vision, grouping refers to the way in which elements in the visual world combine or stay separate from one another in our mental image of the world. Grouping is partly an automatic process, which means that much of it happens rapidly in our brains and without our conscious awareness. It has been described simply as the problem of “what goes with what” in our visual field. Hermann von Helmholtz, the nineteenth-century scientist who taught us much of what we now accept as the foundations of auditory science, described it as an unconscious process that involved inferencing, or logical deductions about what objects in the world are likely to go together based on a number of features or attributes of the objects.

If you’re standing on a mountaintop overlooking a varied landscape, you might describe seeing two or three other mountains, a lake, a valley, a fertile plain, and a forest. Although the forest is composed of hundreds or thousands of trees, the trees form a perceptual group, distinct from other things we see, not necessarily because of our knowledge of forests, but because the trees share similar properties of shape, size, and color—at least when they stand in opposition to fertile plains, lakes, and mountains. But if you’re in the center of a forest with a mixture of alder trees and pines, the smooth white bark of the alders will cause them to “pop out” as a separate group from the craggy dark-barked pines. If I put you in front of one tree and ask you what you see, you might start to focus on details of that tree: bark, branches, leaves (or needles), insects, and moss. When looking at a lawn, most of us don’t typically see individual blades of grass, although we can if we focus our attention on them. Grouping is a hierarchical process and the way in which our brains form

perceptual groups is a function of a great many factors. Some grouping factors are intrinsic to the objects themselves—shape, color, symmetry, contrast, and principles that address the continuity of lines and edges of the object. Other grouping factors are psychological, that is, mind based, such as what we're consciously trying to pay attention to, what memories we have of this or similar objects, and what our expectations are about how objects should go together.

Sounds group too. This is to say that while some group with one another, others segregate from each other. Most people can't isolate the sound of one of the violins in an orchestra from the others, or one of the trumpets from the others—they form a group. In fact, the entire orchestra can form a single perceptual group—called a stream in Bregman's terminology—depending on the context. If you're at an outdoor concert with several ensembles playing at once, the sounds of the orchestra in front of you will cohere into a single auditory entity, separate from the other orchestras behind you and off to the side. Through an act of volition (attention) you can then focus on just the violins of the orchestra in front of you, just as you can follow a conversation with the person next to you in a crowded room full of conversations.

One case of auditory grouping is the way that the many different sounds emanating from a single musical instrument cohere into a percept of a single instrument. We don't hear the individual harmonics of an oboe or of a trumpet, we hear an oboe or we hear a trumpet. This is all the more remarkable if you imagine an oboe and a trumpet playing at the same time. Our brains are capable of analyzing the dozens of different frequencies reaching our ears, and putting them together in just the right way. We don't have the impression of dozens of disembodied harmonics, nor do we hear just a single hybrid instrument. Rather, our brains construct for us separate mental images of an oboe and of a trumpet, and also of the sound of the two of them playing together—the basis for our appreciation of timbral combinations in music. This is what Pierce was talking about when he marveled at the timbres of rock music—the sounds that an electric bass and an electric guitar made when they were playing together—two instruments, perfectly distinguishable from one

another, and yet simultaneously creating a new sonic combination that can be heard, discussed, and remembered.

Our auditory system exploits the harmonic series in grouping sounds together. Our brains coevolved in a world in which many of the sounds that our species encountered—over the tens of thousands of years of evolutionary history—shared certain acoustical properties with one another, including the harmonic series as we now understand it. Through this process of “unconscious inference” (as von Helmholtz called it), our brains assume that it is highly unlikely that several different sound sources are present, each producing a single component of the harmonic series. Rather, our brains use the “likelihood principle” that it must be a single object producing these harmonic components. All of us can make these inferences, even those of us who can't identify or name the instrument “oboe” as distinct, from, say, a clarinet or bassoon, or even a violin. But just as people who don't know the names of the notes in the scale can still tell when two different notes are being played as opposed to the same notes, nearly all of us—even lacking a knowledge of the names of musical instruments—can tell when there are two different instruments playing. The way in which we use the harmonic series to group sounds goes a long way toward explaining why we hear a trumpet rather than the individual overtones that impinge on our ears—they group together like blades of grass that give us the impression of “lawn.” It also explains how we can distinguish a trumpet from an oboe when they're each playing different notes—different fundamental frequencies give rise to a different set of overtones, and our brains are able to effortlessly figure out what goes with what, in a computational process that resembles what a computer might do. But it doesn't explain how we might be able to distinguish a trumpet from an oboe when they're playing the same note, because then the overtones are very nearly the same in frequency (although with different amplitudes characteristic of the instrument). For that, the auditory system relies on a principle of simultaneous onsets. Sounds that begin together—at the same instant in time—are perceived as going together, in the grouping sense. And it has been known since the time Wilhelm Wundt set up the first psychological laboratory in the 1870s

that our auditory system is exquisitely sensitive to what constitutes simultaneous in this sense, being able to detect differences in onset times as short as a few milliseconds.

So when a trumpet and an oboe are playing the same note at the same time, our auditory system is able to figure out that two different instruments are playing because the full sound spectrum—the overtone series—for one instrument begins perhaps a few thousandths of a second before the sound spectrum for the other. This is what is meant by a grouping process that not only integrates sounds into a single object, but segregates them into different objects.

This principle of simultaneous onsets can be thought of more generally as a principle of temporal positioning. We group all the sounds that the orchestra is making now as opposed to those it will make tomorrow night. Time is a factor in auditory grouping. Timbre is another, and this is what makes it so difficult to distinguish one violin from several that are all playing at once, although expert musicians and conductors can train themselves to do this. Spatial location is a grouping principle, as our ears tend to group together sounds that come from the same relative position in space. We are not very sensitive to location in the up-down plane, but we are very sensitive to position in the left-right plane and somewhat sensitive to distance in the forward-back plane. Our auditory system assumes that sounds coming from a distinct location in space are probably part of the same object-in-the-world. This is one of the explanations for why we can follow a conversation in a crowded room relatively easily—our brains are using the cues of spatial location of the person we're conversing with to filter out other conversations. It also helps that the person we're speaking to has a unique timbre—the sound of his voice—that works as an additional grouping cue.

Amplitude also affects grouping. Sounds of a similar loudness group together, which is how we are able to follow the different melodies in Mozart's divertimenti for woodwinds. The timbres are all very similar, but some instruments are playing louder than others, creating different streams in our brains. It is as though a filter or sieve takes the sound of

the woodwind ensemble and separates it out into different parts depending on what part of the loudness scale they are playing in.

Frequency, or pitch, is a strong and fundamental consideration in grouping. If you've ever heard a Bach flute partita, there are typically moments when some flute notes seem to "pop out" and separate themselves from one another, particularly when the flautist is playing a rapid passage—the auditory equivalent of a "Where's Waldo?" picture. Bach knew about the ability of large frequency differences to segregate sounds from one another—to block or inhibit grouping—and he wrote parts that included large leaps in pitch of a perfect fifth or more. The high notes, alternating with a succession of lower-pitched notes, create a separate stream and give the listener the illusion of two flutes playing when there is only one. We hear the same thing in many of the violin sonatas by Locatelli. Yodelers can accomplish the same effect with their voices, by combining pitch and timbral cues; when a male yodeler jumps into his falsetto register, he is creating both a distinct timbre and, typically, a large jump in pitch, causing the higher notes to again separate out into a distinct, perceptual stream, giving the illusion of two people singing interleaved parts.

We now know that the neurobiological subsystems for the different attributes of sound that I've described separate early on, at low levels of the brain. This suggests that grouping is carried out by general mechanisms working somewhat independently of one another. But it is also clear that the attributes work with or against each other when they combine in particular ways, and we also know that experience and attention can have an influence on grouping, suggesting that portions of the grouping process are under conscious, cognitive control. The ways in which conscious and unconscious processes work together—and the brain mechanisms that underlie them—are still being debated, but we've come a long way toward understanding them in the last ten years. We've finally gotten to the point where we can pinpoint specific areas of the brain that are involved in particular aspects of music processing. We even think we know which part of the brain causes you to pay attention to things.

How are thoughts formed? Are memories “stored” in a particular part of the brain? Why do songs sometimes get stuck in your head and you can’t get them out? Does your brain take some sick pleasure in slowly driving you crazy with inane commercial jingles? I take up these and other ideas in the coming chapters.

3. Behind the Curtain

Music and the Mind Machine

For cognitive scientists, the word *mind* refers to that part of each of us that embodies our thoughts, hopes, desires, memories, beliefs, and experiences. The brain, on the other hand, is an organ of the body, a collection of cells and water, chemicals and blood vessels, that resides in the skull. Activity in the brain gives rise to the contents of the mind. Cognitive scientists sometimes make the analogy that the brain is like a computer’s CPU, or hardware, while the mind is like the programs or software running on the CPU. (If only that were literally true and we could just run out to buy a memory upgrade.) Different programs can run on what is essentially the same hardware—different minds can arise from very similar brains.

Western culture has inherited a tradition of dualism from René Descartes, who wrote that the mind and the brain are two entirely separate things. Dualists assert that the mind preexisted, before you were born, and that the brain is not the seat of thought—rather, it is merely an instrument of the mind, helping to implement the mind’s will, move muscles, and maintain homeostasis in the body. To most of us, it certainly feels as though our minds are something unique and distinctive, separate from just a bunch of neurochemical processes. We have a feeling of what it is like to be me, what it is like to be me reading a book, and what it is