

1. What Is Music?

From Pitch to Timbre

What is music? To many, "music" can only mean the great masters—Beethoven, Debussy, and Mozart. To others, "music" is Busta Rhymes, Dr. Dre, and Moby. To one of my saxophone teachers at Berklee College of Music—and to legions of "traditional jazz" aficionados—anything made before 1940 or after 1960 isn't *really* music at all. I had friends when I was a kid in the sixties who used to come over to my house to listen to the Monkees because their parents forbade them to listen to anything but classical music, and others whose parents would only let them listen to and sing religious hymns, in both cases fearing the "dangerous rhythms" of rock and roll. When Bob Dylan dared to play an electric guitar at the Newport Folk Festival in 1965, people walked out and many of those who stayed, booed. The Catholic Church banned music that contained polyphony (more than one musical part playing at a time), fearing that it would cause people to doubt the unity of God. The church also banned the musical interval of an augmented fourth, the distance between C and F-sharp and also known as a tritone (the interval in Leonard Bernstein's *West Side Story* when Tony sings the name "Maria"). This interval was considered so dissonant that it must have been the work of Lucifer, and so the church named it *Diabolus in musica*. It was pitch that had the medieval church in an uproar. And it was timbre that got Dylan booed. It was the latent African rhythms in rock that frightened white suburban parents, perhaps fearful that the beat would induce a permanent, mind-altering trance in their innocent children. What are rhythm, pitch, and timbre—are they merely ways of describing different mechanical aspects of a song, or do they have a deeper, neurological basis? Are all of these elements necessary?

The music of avant-garde composers such as Francis Dhomont, Robert Normandeau, or Pierre Schaeffer stretches the bounds of what most of us think music is. Going beyond the use of melody and harmony, and even beyond the use of instruments, these composers use recordings of found objects in the world such as jackhammers, trains, and waterfalls. They edit the recordings, play with their pitch, and ultimately combine them into an organized collage of sound with the same type of emotional trajectory—the same tension and release—as traditional music. Composers in this tradition are like the painters who stepped outside of the boundaries of representational and realistic art—the cubists, the

Dadaists, many of the modern painters from Picasso to Kandinsky to Mondrian.

What do the music of Bach, Depeche Mode, and John Cage fundamentally have in common? On the most basic level, what distinguishes Busta Rhymes's "What's It Gonna Be?!" or Beethoven's "Pathétique" Sonata from, say, the collection of sounds you'd hear standing in the middle of Times Square, or those you'd hear deep in a rainforest? As the composer Edgard Varèse famously defined it, "Music is organized sound."

This book drives at a neuropsychological perspective on how music affects our brains, our minds, our thoughts, and our spirit. But first, it is helpful to examine what music is made of. What are the fundamental building blocks of music? And how, when organized, do they give rise to music? The basic elements of any sound are loudness, pitch, contour, duration (or rhythm), tempo, timbre, spatial location, and reverberation. Our brains organize these fundamental perceptual attributes into higher-level concepts—just as a painter arranges lines into forms—and these include meter, harmony, and melody. When we listen to music, we are actually perceiving multiple attributes or "dimensions."

Before getting to the brain basis of all this, I'd like to take this chapter to define the musical terms and quickly review some basic ideas in music theory, and illustrate them with musical examples. (Musicians may want to skip or skim this chapter.) First here is a brief summary of the main terms.

~ *Pitch* is a purely psychological construct, related both to the actual frequency of a particular tone and to its relative position in the musical scale. It provides the answer to the question "What note is that?" ("It's a C-sharp.") I'll define frequency and musical scale below. (When a trumpet player blows in his instrument and makes a single sound, he makes what most of us call a *note*, and what scientists call a *tone*. The two terms, *note* and *note* refer to the same entity in the abstract, but we reserve the word *tone* for what you hear, and the word *note* for what you see written on a musical score.) In the nursery rhymes "Mary Had a Little Lamb" and "Are You Sleeping?" pitch is the *only* thing that varies in the first seven notes—the rhythm stays the same. This demonstrates the power—and fundamentality—of pitch in defining a melody or song.

~ *Rhythm* refers to the durations of a series of notes, and to the way that they group together into units. For example, in the "Alphabet Song" (the same as "Twinkle, Twinkle Little Star") the first six notes of the song are all equal in duration as we sing the names of the letters A B C D E F and then we hold the letter G for twice the duration. Then we're back to the standard duration for H I J K, and then the following four letters are sung with half the

duration, or twice as fast per letter: L M N O and then ending on a held P (leading generations of schoolchildren to spend several early months believing that there was a letter in the English alphabet called *ellenmeno*). In the Beach Boys' song "Barbara Ann," the first seven notes are all sung on the same pitch, with only the rhythm varying. In fact, the seven notes after that are all sung on the same pitch as well (in the melody), as Dean Torrence (of Jan & Dean) is joined by other voices singing other notes (harmony). The Beatles have several songs in which pitch is held constant and only rhythm varies across several notes: the first four notes of "Come Together"; the six notes of "Hard Day's Night" following the lyric "It's been a"; the first six notes of "Something."

~ *Tempo* refers to the overall speed or pace of the piece. If you were tapping your foot, dancing, or marching to the piece, it's how fast or slow these regular movements would be.

~ *Contour* describes the overall shape of a melody, taking into account only the pattern of "up" and "down" (whether a note goes up or down, not the amount by which it goes up or down).

~ *Timbre* (rhymes with *amber*) distinguishes one instrument from another—say, trumpet from piano—when both are playing the same written note. It is a kind of tonal color that is produced in part by overtones from the instrument's vibrations (more on that later). It also describes the way that a single instrument can change sound as it moves across its range—say the warm sound of a trumpet low in its range versus the piercing sound of that same trumpet playing its highest note.

~ *Loudness* is a purely psychological construct that relates (nonlinearly and in poorly understood ways) to how much energy an instrument creates—how much air it displaces—and what an acoustician would call the amplitude of a tone.

~ *Reverberation* refers to the perception of how distant the source is from us in combination with how large a room or hall the music is in; often referred to as "echo" by laypeople, it is the quality that distinguishes the spaciousness of singing in a large concert hall from the sound of singing in your shower. It has an underappreciated role in communicating emotion and creating an overall pleasing sound.

Psychophysicists—scientists who study the ways that the brain interacts with the physical world—have shown that these attributes are *separable*. Each can be varied without altering the others, allowing the scientific study of one at a time. I can change the pitches in a song without changing the rhythm, and I can play a song on a different instrument (changing the timbre) without changing the duration or pitches of the

notes. The difference between *music* and a random or disordered set of sounds has to do with the way these fundamental attributes combine, and the relations that form between them. When these basic elements combine and form relationships with one another in a meaningful way, they give rise to higher-order concepts such as meter, key, melody, and harmony.

~ *Meter* is created by our brains by extracting information from rhythm and loudness cues, and refers to the way in which tones are grouped with one another across time. A waltz meter organizes tones into groups of three, a march into groups of two or four.

~ *Key* has to do with a hierarchy of importance that exists between tones in a musical piece; this hierarchy does not exist in-the-world, it exists only in our minds, as a function of our experiences with a musical style and musical idioms, and mental schemas that all of us develop for understanding music.

~ *Melody* is the main theme of a musical piece, the part you sing along with, the succession of tones that are most salient in your mind. The notion of melody is different across genres. In rock music, there is typically a melody for the verses and a melody for the chorus, and verses are distinguished by a change in lyrics and sometimes by a change in instrumentation. In classical music, the melody is a starting point for the composer to create variations on that theme, which may be used throughout the entire piece in different forms.

~ *Harmony* has to do with relationships between the pitches of different tones, and with tonal contexts that these pitches set up that ultimately lead to expectations for what will come next in a musical piece—expectations that a skillful composer can either meet or violate for artistic and expressive purposes. Harmony can mean simply a parallel melody to the primary one (as when two singers harmonize) or it can refer to a chord progression—the clusters of notes that form a context and background on which the melody rests.

I'll be elaborating on all of these as we go along.

The idea of primitive elements combining to create art, and of the importance of relationships between elements, also exists in visual art and dance. The fundamental elements of visual perception include color (which itself can be decomposed into the three dimensions of hue, saturation, and lightness), brightness, location, texture, and shape. But a painting is more than these—it is not just a line here and another there, or a spot of red in one part of the picture and a patch of blue in another. What makes a set of lines and colors into art is the *relationship* between this line and that one; the way one color or form echoes another in a different part of the canvas. Those dabs of paint and lines become art

when form and flow (the way in which your eye is drawn across the canvas) are created out of lower-level perceptual elements. When they combine harmoniously they give rise to perspective, foreground and background, and ultimately to emotion and other aesthetic attributes. Similarly, dance is not just a raging sea of unrelated bodily movements; the relationship of those movements to one another is what creates integrity and integrality, a coherence and cohesion that the higher levels of our brain process. And as in visual art, music plays on not just what notes are sounded, but which ones are not. Miles Davis famously described his improvisational technique as parallel to the way that Picasso described his use of a canvas: The most critical aspect of the work, both artists said, was not the objects themselves, but the space between objects. In Miles's case, he described the most important part of his solos as the empty space between notes, the "air" that he placed between one note and the next. Knowing precisely when to hit the next note, and allowing the listener time to anticipate it, is a hallmark of Davis's genius. This is particularly apparent in his album *Kind of Blue*.

* * *

To nonmusicians, terms such as *diatonic*, *cadence*, or even *key* and *pitch* can throw up an unnecessary barrier. Musicians and critics sometimes appear to live behind a veil of technical terms that can sound pretentious. How many times have you read a concert review in the newspaper and found you have no idea what the reviewer is saying? "Her sustained *appoggiatura* was flawed by an inability to complete the *roulade*." Or, "I can't believe they modulated to C-sharp minor! How ridiculous!" What we really want to know is whether the music was performed in a way that moved the audience. Whether the singer seemed to inhabit the character she was singing about. You might want the reviewer to compare tonight's performance to that of a previous night or a different ensemble. We're usually interested in the music, not the technical devices that were used. We wouldn't stand for it if a restaurant reviewer started to speculate about the precise temperature at which the chef introduced the lemon juice in a hollandaise sauce, or if a film critic talked about the aperture of the lens that the cinematographer used; we shouldn't stand for it in music either.

Moreover, those who study music—even musicologists and scientists—disagree about what is meant by some of these terms. We employ the term *timbre*, for example, to refer to the overall sound or tonal color of an instrument—that indescribable character that distinguishes a trumpet from a clarinet when they're playing the same written note, or what distinguishes your voice from Brad Pitt's if you're saying the same words. But an inability to agree on a definition has caused the scientific community to take the unusual step of throwing up its hands and defining timbre by what it is not. (The official definition of the Acoustical Society of America is that timbre is everything about a sound that is not loudness

or pitch. So much for scientific precision!)

What is pitch and where does it come from? This simple question has generated hundreds of scientific articles and thousands of experiments. Almost all of us, even without musical training, can tell if a singer is offkey; we might not be able to say whether she is sharp or flat, or by how much, but after the age of five, most humans have as well a refined ability to detect tones that are out of tune as to discriminate a question from an accusation (in English, a rising pitch indicates a question, a straight or slightly falling pitch indicates an accusation). This comes from an interaction between our exposure to music and the physics of sound. What we call pitch is related to the frequency or rate of vibration of a string, column of air, or other physical source. If a string is vibrating so that it moves back and forth sixty times in one second, we say that it has a frequency of sixty cycles per second. The unit of measurement, cycles per second, is often called Hertz (abbreviated Hz) after Heinrich Hertz, the German theoretical physicist who was the first to transmit radio waves (a dyed-in-the-wool theoretician, when asked what practical use radio waves might have, he reportedly shrugged, "None"). If you were to try to mimic the sound of a fire engine siren, your voice would sweep through different pitches, or frequencies (as the tension in your vocal folds changes), some "low" and some "high."

Keys on the left of the piano keyboard strike longer, thicker strings that vibrate at a relatively slow rate. Keys to the right strike shorter, thinner strings that vibrate at a higher rate. The vibration of these strings displaces air molecules, and causes them to vibrate at the same rate—with the same frequency as the string. These vibrating air molecules are what reach our eardrum, and they cause our eardrum to wiggle in and out at the same frequency. The only information that our brains get about the pitch of sound comes from that wiggling in and out of our eardrum; our inner ear and our brain have to analyze the motion of the eardrum in order to figure out what vibrations out-there-in-the-world caused the eardrum to move that way. Although I said that air molecules vibrate, other molecules will too—we can hear music under water or in other fluids if the water (or other fluid) molecules are caused to vibrate. But in the vacuum of space, with no molecules to vibrate, there is no sound. (The next time you're watching *Star Trek* and hear the roar of the engines in space, you'll have some good Trekkie Trivia to share.)

By convention, when we press keys nearer to the left of the keyboard, we say that they are "low" pitch sounds, and ones near the right side of the keyboard are "high" pitch. That is, what we call "low" are those sounds that vibrate slowly, and are closer (in vibration frequency) to the sound of a large dog barking. What we call "high" are those sounds that vibrate rapidly, and are closer to what a small yip-yip dog might make. But even these terms *high* and *low* are culturally relative—the Greeks talked about sounds in the opposite way because the stringed instruments they built tended to be oriented vertically. Shorter strings or pipe organ tubes had

their tops closer to the ground, so these were called the “low” notes (as in “low to the ground,”) and the longer strings and tubes—reaching up toward Zeus and Apollo—were called the “high” notes. *Low* and *high*—just like *left* and *right*—are effectively arbitrary terms that ultimately have to be memorized. Some writers have argued that “high” and “low” are intuitive labels, noting that what we call high-pitched sounds come from birds (who are high up in trees or in the sky) and what we call low-pitched sounds often come from large, close-to-the-ground mammals such as bears or the low sounds of an earthquake. But this is not convincing, since low sounds also come from up high (think of thunder) and high sounds can come from down low (crickets and squirrels, leaves being crushed underfoot).

As a first definition of *pitch*, let’s say it is that quality that primarily distinguishes the sound that is associated with pressing one piano key versus another.

Pressing a piano key causes a hammer to strike one or more strings inside the piano. Striking a string displaces it, stretching it a bit, and its inherent resiliency causes it to return toward its original position. But it overshoots that original position, going too far in the opposite direction, and then attempts to return to its original position again, overshooting it again, and in this way it oscillates back and forth. Each oscillation covers less distance, and, in time, the string stops moving altogether. This is why the sound you hear when you press a piano key gets softer until it trails off into nothing. The distance that the string covers with each oscillation back and forth is translated by our brains into loudness; the rate at which it oscillates is translated into pitch. The farther the string travels, the louder the sound seems to us; when it is barely traveling at all, the sound seems soft. Although it might seem counterintuitive, the distance traveled and the rate of oscillation are independent. A string can vibrate very quickly and traverse either a great distance or a small one. The distance it traverses is related to how hard we hit it—this corresponds to our intuition that hitting something harder makes a louder sound. The rate at which the string vibrates is principally affected by its size and how tightly strung it is, not by how hard it was struck.

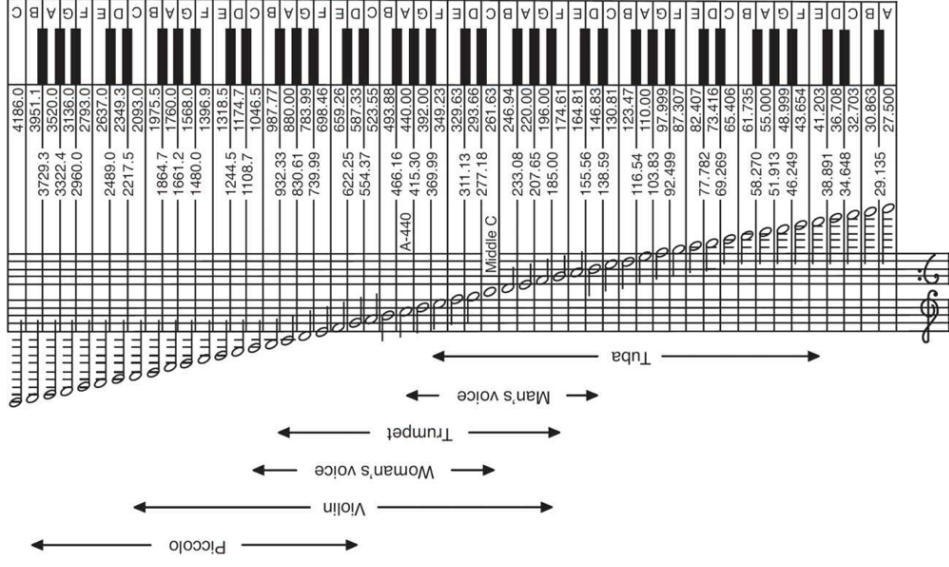
It might seem as though we should simply say that pitch is the same as frequency, that is, the frequency of vibration of air molecules. This is almost true. Mapping the physical world onto the mental world is seldom so straightforward, as we’ll see later. However, for most musical sounds, pitch and frequency are closely related.

The word *pitch* refers to the mental representation an organism has of the fundamental frequency of a sound. That is, pitch is a purely psychological phenomenon related to the frequency of vibrating air molecules. By “psychological,” I mean that it is entirely in our heads, not in the world-out-there; it is the end product of a chain of mental events that gives rise to an entirely subjective, internal mental representation or quality. Sound waves—molecules of air vibrating at various frequencies—

do not themselves have pitch. Their motion and oscillations can be measured, but it takes a human (or animal) brain to map them to that internal quality we call pitch.

We perceive color in a similar way, and it was Isaac Newton who first realized this. (Newton, of course, is known as the discoverer of the theory of gravity, and the inventor, along with Leibniz, of calculus. Like Einstein, Newton was a very poor student, and his teachers often complained of his inattentiveness.)

Newton was the first to point out that light is colorless, and that consequently color has to occur inside our brains. He wrote, “The waves themselves are not colored.” Since his time, we have learned that light waves are characterized by different frequencies of oscillation, and when they impinge on the retina of an observer, they set off a chain of neurochemical events, the end product of which is an internal mental image that we call color. The essential point here is: What we perceive as color is not made up of color. Although an apple may appear red, its atoms are not themselves red. And similarly, as the philosopher Daniel Dennett points out, heat is not made up of tiny hot things.



A bowl of pudding only has taste when I put it in my mouth—when it is in contact with my tongue. It doesn't have taste or flavor sitting in my fridge, only the potential. Similarly, the walls in my kitchen are not “white” when I leave the room. They still have paint on them, of course, but *color* only occurs when they interact with my eyes.

Sound waves impinge on the eardrums and pinnae (the fleshy parts of your ear), setting off a chain of mechanical and neurochemical events, the end product of which is an internal mental image we call pitch. If a tree falls in a forest and no one is there to hear it, does it make a sound? (The question was first posed by the Irish philosopher George Berkeley.) Simply, no—sound is a mental image created by the brain in response to vibrating molecules. Similarly, there can be no pitch without a human or animal present. A suitable measuring device can register the frequency made by the tree falling, but truly it is not pitch unless and until it is heard.

No animal can hear a pitch for every frequency that exists, just as the colors that we actually see are only a small portion of the entire electromagnetic spectrum. Sound can theoretically be heard for vibrations from just over 0 cycles per second up to 100,000 cycles per second or more, but each animal hears only a subset of the possible sounds. Humans who are not suffering from any kind of hearing loss can usually hear sounds from 20 Hz to 20,000 Hz. The pitches at the low end sound like an indistinct rumble or shaking—this is the sound we hear when a truck goes by outside the window (its engine is creating sound around 20 Hz) or when a tricked-out car with a fancy sound system has the subwoofers cranked up really loud. Some frequencies—those below 20 Hz—are inaudible to humans because the physiological properties of our ears aren't sensitive to them. The beats we hear on 50 Cents' "In da Club" or N.W.A.'s "Express Yourself" are near the low end of our range of hearing; the ending of "A Day in Life" on the CD of the Beatles' *Sgt. Pepper's Lonely Hearts Club Band* has a few seconds of sound at 15 KHz, inaudible to most adults over 40! (If the Beatles believed to never trust anyone over 40, this may have been their test, but Lennon reportedly just wanted something to make people's dogs perk up.)

The range of human hearing is generally 20 Hz to 20,000 Hz, but this doesn't mean that the range of human pitch perception is the same; although we can hear sounds in this entire range, they don't all sound musical; that is, we can't unambiguously assign a pitch to the entire range. By analogy, colors at the infrared and ultraviolet ends of the spectrum lack definition compared to the colors closer to the middle. The figure on [page 23](#) shows the range of musical instruments, and the frequency associated with them. The sound of the average male speaking voice is around 110 Hz, and the average female speaking voice is around 220 Hz. The hum of fluorescent lights or from faulty wiring is 60 Hz (in North America; in Europe and countries with a different voltage/current standard, it can be 50 Hz). The sound that a singer hits when she causes a glass to break might be 1000 Hz. The glass breaks because it, like all physical objects, has a natural and inherent vibration frequency. You can hear this by flicking your finger against its sides or, if it's crystal, by running your wet finger around the rim of the glass in a circular motion. When the singer hits just the right frequency—the resonant frequency of

the glass—it causes the molecules of the glass to vibrate at their natural rate, and they vibrate themselves apart.

A standard piano has eighty-eight keys. Very rarely, pianos can have a few extra ones at the bottom and electronic pianos, organs, and synthesizers can have as few as twelve or twenty-four keys, but these are special cases. The lowest note on a standard piano vibrates with a frequency of 27.5 Hz. Interestingly, this is about the same rate of motion that constitutes an important threshold in visual perception. A sequence of still photographs—slides—displayed at or about this rate of presentation will give the illusion of motion. "Motion pictures" are a sequence of still images presented at a rate (twenty-four frames per second) that exceeds the temporal resolving properties of the human visual system. In 35 mm film projection, each image is presented for $\approx 1/48$ th of a second, alternating with a black frame of roughly equal duration as the lens is blocked between successive still images. We perceive smooth, continuous motion when in fact there is no such thing actually being shown to us. (Old-timey movies seem to flicker because their frame rate, at 16–18 fps) was too low, and our visual system picked up on the discontinuities.) When molecules vibrate at around this speed we hear something that sounds like a continuous tone. If you put playing cards in the spokes of your bicycle wheel when you were a kid, you demonstrated to yourself a related principle: At slow speeds, you simply hear the click-click-click of the card hitting the spokes. But above a certain speed, the clicks run together and create a buzz, a tone you can actually hum along with; a pitch.

When that lowest note on the piano plays, and vibrates at 27.5 Hz, to most people it lacks the distinct pitch of sounds toward the middle of the keyboard. At the lowest and the highest ends of the piano keyboard, the notes sound fuzzy to many people with respect to their pitch. Composers know this, and they either use these notes or avoid them depending on what they are trying to accomplish compositionally and emotionally. Sounds with frequencies above the highest note on the piano keyboard, around 6000 Hz and more, sound like a high-pitched whistling to most people. Above 20,000 Hz most humans don't hear a thing, and by the age of sixty, most adults can't hear much above 15,000 Hz or so due to a stiffening of the hair cells in the inner ear. So when we talk about the range of musical notes, or that restricted part of the piano keyboard that conveys the strongest sense of pitch, we are talking about roughly three quarters of the notes on the piano keyboard, between about 55 Hz and 2000 Hz.

Pitch is one of the primary means by which musical emotion is conveyed. Mood, excitement, calm, romance, and danger are signaled by a number of factors, but pitch is among the most decisive. A single high note can convey excitement, a single low note sadness. When notes are strung together, we get more powerful and more nuanced musical statements. Melodies are defined by the pattern or relation of successive pitches across

time; most people have no trouble recognizing a melody that is played in a higher or lower key than they've heard it in before. In fact, many melodies do not have a "correct" starting pitch, they just float freely in space, starting anywhere. "Happy Birthday" is an example of this. One way to think about a melody, then, is as an *abstract prototype* that is derived from specific combinations of key, tempo, instrumentation, and so on. A cognitive psychologist would say that a melody is an auditory object that maintains its identity in spite of transformations, just as a chair maintains its identity when you move it to the other side of the room, turn it upside down, or paint it red. So, for example, if you hear a song played louder than you are accustomed to, you still identify it as the same song. The same holds for changes in the absolute pitch values of the song, which can be changed so long as the relative distances between them remain the same.

The notion of relative pitch values is seen readily in the way that we speak. When you ask someone a question, your voice naturally rises in intonation at the end of the sentence, signaling that you are asking. But you don't try to make the rise in your voice match a specific pitch. It is enough that you end the sentence somewhat higher in pitch than you began it. This is a convention in English (though not in all languages—we have to learn it), and is known in linguistics as a prosodic cue. There are similar conventions for music written in the Western tradition. Certain sequences of pitches evoke calm, others, excitement. The slow, predominantly step-wise downward motion of the melody in Grieg's "Peer Gynt Suite No. 1, Morning Mood" conveys peacefulness; but in "Anitra's Dance" from the same suite, the chromatic, ascending lines (with occasional and playfully descending intervals on the way up) we sense more activity and movement. The brain basis for this is primarily based on learning, just as we learn that a rising intonation indicates a question. All of us have the innate capacity to learn the linguistic and musical distinctions of whatever culture we are born into, and experience with the music of that culture shapes our neural pathways so that we ultimately internalize a set of rules common to that musical tradition.

Different instruments use different parts of the range of available pitches. The piano has the largest range of any instrument, as you can see from the previous illustration. The other instruments each use a subset of the available pitches, and this influences the ways that instruments are used to communicate emotion. The piccolo, with its high-pitched, shrill, and birdlike sound, tends to evoke flighty, happy moods regardless of the notes it's playing. Because of this, composers tend to use the piccolo for happy music, or rousing music, as in a Sousa march. Similarly, in *Peter and the Wolf*, Prokofiev uses the flute to represent the bird, and the French horn to indicate the wolf. The characters' individuality in *Peter and the Wolf* is expressed in the timbres of different instruments and each has a leitmotif—an associated melodic phrase or figure that accompanies the reappearance of an idea, person, or situation. (This is especially true of

Wagnerian music drama.) A composer who picks so-called sad pitch sequences would only give these to the piccolo if he were trying to be ironic. The lumbering, deep sounds of the tuba or double bass are often used to evoke solemnity, gravity, or weight.

How many different pitches are there? Because pitch comes from a continuum—the vibration frequencies of molecules—there are technically an infinite number of pitches: For every pair of frequencies you mention, I could always come up with one between them, and a theoretically different pitch would exist. But not every change in frequency gives rise to a noticeable difference in pitch, just as adding a grain of sand to your backpack will not change the weight perceptibly. So not all frequency changes are musically useful. People differ in their ability to detect small changes in frequency; training can help, but generally speaking, most cultures don't use distances much smaller than a semitone as the basis for their music, and most people can't reliably detect changes smaller than about one-tenth of a semitone.

The ability to detect differences in pitch is based on physiology, and varies from one animal to another. How is it that we humans can tell pitches apart? The basilar membrane of the inner ear contains hair cells that are frequency selective, firing only in response to a certain band of frequencies. These are stretched out across the membrane from low frequencies to high; low-frequency sounds excite hair cells on one end of the basilar membrane, medium frequency sounds excite the hair cells in the middle, and high-frequency sounds excite them at the other end. We can think of the membrane as containing a map of different pitches very much like a piano keyboard superimposed on it. Because the different tones are spread out across the surface topography of the membrane, this is called a tonotopic map.

After sounds enter the ear, they pass by the basilar membrane, where certain hair cells fire, depending on the frequency of the sounds. The membrane acts like a motion-detector lamp you might have in your garden; activity in a certain part of the membrane causes it to send an electrical signal on up to the auditory cortex. The auditory cortex also has a tonotopic map, with low to high tones stretched out across the cortical surface. In this sense, the brain also contains a "map" of different pitches, and different areas of the brain respond to different pitches. Pitch is so important that the brain represents it directly: unlike almost any other musical attribute, we could place electrodes in the brain and be able to determine what pitches were being played to a person just by looking at the brain activity. And although music is based on pitch relations rather than absolute pitch values, it is, paradoxically, these absolute pitch values that the brain is paying attention to throughout its different stages of processing.

This direct mapping of pitch is so important, it bears repeating. If I put electrodes in your visual cortex (the part of the brain at the back of the head, concerned with seeing), and I then showed you a red tomato, there is

no group of neurons that will cause my electrodes to turn red. But if I put electrodes in your auditory cortex and play a pure tone in your ears at 440 Hz, there are neurons in your auditory cortex that will fire at precisely that frequency, causing the electrode to emit electrical activity at 440 Hz—for pitch, what goes into the ear comes out of the brain!

* * *

A scale is just a subset of the theoretically infinite number of pitches, and every culture selects these based on historical tradition or somewhat arbitrarily. The specific pitches chosen are then anointed as being part of that musical system. These are the letters that you see in the figure above. The names "A," "B," "C," and so on are arbitrary labels that we associate with particular frequencies. In Western music—music of the European tradition—these pitches are the only "legal" pitches; most instruments are designed to play these pitches and not others. (Instruments like the trombone and cello are an exception, because they can slide between notes; trombonists, cellists, violinists, etc., spend a lot of time learning how to hear and produce the precise frequencies required to play each of the legal notes.) Sounds in between are considered mistakes ("out of tune") unless they're used for expressive intonation (intentionally playing something out of tune, briefly, to add emotional tension) or in passing from one legal tone to another.

Tuning refers to the precise relationship between the frequency of a tone being played and a standard, or between two or more tones being played together. Orchestral musicians "tuning up" before a performance are synchronizing their instruments (which naturally drift in their tuning as the wood, metal, strings, and other materials expand and contract with changes in temperature and humidity) to a standard frequency, or occasionally not to a standard but to each other. Expert musicians often alter the frequency of tones while they're playing for expressive purposes (except, of course, on fixed-pitch instruments such as keyboards and xylophones); sounding a note slightly lower or higher than its nominal value can impart emotion when done skillfully. Expert musicians playing together in ensembles will also alter the pitch of tones they play to bring them more in tune with the tones being played by the other musicians, should one or more musicians drift away from standard tuning during the performance.

The note names in Western music run from A to G, or, in an alternative system, as Do - re - mi - fa - sol - la - ti - do (the alternate system is used as lyrics to the Rodgers and Hammerstein song "Do-Re-Mi" from *The Sound of Music*: "Do, a deer, a female deer, Re, a drop of golden sun . . ."). As frequencies get higher, so do the letter names; B has a higher frequency than A (and hence a higher pitch) and C has a higher frequency than either A or B. After G, the note names start all over again at A. Notes with the same name have frequencies that are double (or half) the frequencies of each other. One of the several notes we call A has a frequency of 110 Hz.

The note with half that frequency—55 Hz—is also an A and the note with twice 110 Hz—220 Hz—is an A as well. If we keep doubling the frequencies we get more As at 440 Hz, 880 Hz, 1760 Hz, and so on.

Here is a fundamental quality of music. Note names repeat because of a perceptual phenomenon that corresponds to the doubling and halving of frequencies. When we double or halve a frequency, we end up with a note that sounds remarkably similar to the one we started out with. This relationship, a frequency ratio of 2:1 or 1:2, is called the octave. It is so important that, in spite of the large differences that exist between musical cultures—between Indian, Balinese, European, Middle Eastern, Chinese, and so on—every culture we know of has the octave as the basis for its music, even if it has little else in common with other musical traditions. This phenomenon leads to the notion of circularity in pitch perception, and is similar to circularity in colors. Although red and violet fall at opposite ends of the continuum of visible frequencies of electromagnetic energy, we see them as perceptually similar. The same is true in music, and music is often described as having two dimensions, one that accounts for tones going up in frequency (and sounding higher and higher) and another that accounts for the perceptual sense that we've come back home again each time we double a tone's frequency.

When men and women speak in unison, their voices are normally an octave apart, even if they try to speak the exact same pitches. Children generally speak an octave or two higher than adults. The first two notes of the Harold Arlen melody "Over the Rainbow" (from the movie *The Wizard of Oz*) make an octave. In "Hot Fun in the Summertime" by Sly and the Family Stone, Sly and his backup singers are singing in octaves during the first line of the verse "End of the spring and here she comes back." As we increase frequencies by playing the successive notes on an instrument, there is a very strong perceptual sense that when we reach a doubling of frequency, we have come "home" again. The octave is so basic that even some animal species—monkeys and cats, for example—show octave equivalence, the ability to treat as similar, the way that humans do, tones separated by this amount.

An interval is the distance between two tones. The octave in Western music is subdivided into twelve (logarithmically) equally spaced tones. The intervallic distance between A and B (or between "do" and "re") is called a whole step or a tone. (This latter term is confusing, since we call any musical sound a tone; I'll use the term *whole step* to avoid ambiguity.) The smallest division in our Western scale system cuts a whole step perceptually in half: the half step, or semitone, which is one-twelfth of an octave. (I'll use the word *semitone* because it is more common, and because there is no ambiguity about what it means.)

Intervals are the basis of melody, much more so than the actual pitches of notes; melody processing is relational, not absolute, meaning that we define a melody by its intervals, not the actual notes used to create them. Four semitones always create the interval known as a major third

regardless of whether the first note is an A or a G# or any other note. See the table of the intervals as they're known in our (Western) musical system.

The table could continue on: Thirteen semitones is a minor ninth, fourteen semitones is a major ninth, etc., but these names are typically used only in more advanced discussions. The intervals of the perfect fourth and perfect fifth are so called because they sound particularly pleasing to many people, and since the ancient Greeks, this particular feature of the scale is at the heart of all music. (There is no "imperfect fifth," this is just the name we give the interval.) Ignore the perfect fourth and fifth or use them in every phrase, they have been the backbone of music for at least five thousand years.

Although the areas of the brain that respond to individual pitches have been mapped, we have not yet been able to find the neurological basis for the encoding of pitch relations; we know which part of the cortex is involved in listening to the notes C and E, for example, and for F and A, but we do not know how or why both intervals are perceived as a major third, or the neural circuits that create this perceptual equivalency. These relations must be extracted by computational processes in the brain that remain poorly understood.

<i>Distance in semitones</i>	<i>Interval name</i>
0	unison
1	minor second
2	major second
3	minor third
4	major third
5	perfect fourth
6	augmented fourth, diminished fifth, or tritone
7	perfect fifth
8	minor sixth
9	major sixth
10	minor seventh
11	major seventh
12	octave

If there are twelve named notes within an octave, why are there only seven letters (or do-re-mi syllables)? After centuries of being forced to eat in the servants' quarters and to use the back entrance of the castle, this may just be an invention by musicians to make nonmusicians feel inadequate. The additional five notes have compound names, such as **F_b**, pronounced "E-flat") and **F#** (pronounced "F-sharp"). There is no reason for the system to be so complicated, but it is what we're stuck with.

The system is a bit clearer looking at the piano keyboard. A piano has white keys and black keys spaced out in an uneven arrangement—sometimes two white keys are adjacent, sometimes they have a black key between them. Whether the keys are white or black, the perceptual distance from one adjacent key to the next always makes a semitone, and a distance of two keys is always a whole step. This applies to many Western instruments; the distance between one fret on a guitar and the next is also a semitone, and pressing or lifting adjacent keys on woodwind instruments (such as the clarinet or oboe) typically changes the pitch by a semitone.

The white keys are named A, B, C, D, E, F, and G. The notes between—the black keys—are the ones with compound names. The note between A and B is called either A-sharp or B-flat, and in all but formal music theoretic discussions, the two terms are interchangeable. (In fact, this note could also be referred to as C double-flat, and similarly, A could be called G double-sharp, but this is an even more theoretical usage.) Sharp means high, and flat means low. B-flat is the note one semitone lower than B; A-sharp is the note one semitone higher than A. In the parallel do-re-mi system, unique syllables mark these other tones: di and ra indicate the tone between do and re, for example.

The notes with compound names are not in any way second-class musical citizens. They are just as important, and in some songs and some scales they are used exclusively. For example, the main accompaniment to "Superstition" by Stevie Wonder is played on only the black keys of the keyboard. The twelve tones taken together, plus their repeating cousins one or more octaves apart, are the basic building blocks for melody, for all the songs in our culture. Every song you know, from "Deck the Halls" to "Hotel California," from "Ba Ba Black Sheep" to the theme from *Sex and the City*, is made up from a combination of these twelve tones and their octaves.

To add to the confusion, musicians also use the terms *sharp* and *flat* to indicate if someone is playing out of tune; if the musician plays the tone a bit too high (but not so high as to make the next note in the scale) we say that the tone being played is sharp, and if the musician plays the tone too low we say that the tone is flat. Of course, a musician can be only slightly off and nobody would notice. But when the musician is off by a relatively large amount—say one quarter to one half the distance between the note she was trying to play and the next one—most of us can usually detect this and it sounds off. This is especially apparent when there is more than one instrument playing, and the out-of-tune tone we are hearing clashes with in-tune tones being played simultaneously by other musicians.

The names of pitches are associated with particular frequency values. Our current system is called A440 because the note we call A that is in the middle of the piano keyboard has been fixed to have a frequency of 440 Hz. This is entirely arbitrary. We could fix A at any frequency, such as 439, 444, 424, or 314.159; different standards were used in the time of Mozart

than today. Some people claim that the precise frequencies affect the overall sound of a musical piece and the sound of instruments. Led Zeppelin often tuned their instruments away from the modern A440 standard to give their music an uncommon sound, and perhaps to link it with the European children's folk songs that inspired many of their compositions. Many purists insist on hearing baroque music on period instruments, both because the instruments have a different sound and because they are designed to play the music in its original tuning standard, something that purists deem important.

We can fix pitches anywhere we want because what defines music is a set of pitch relations. The specific frequencies for notes may be arbitrary, but the distance from one frequency to the next—and hence from one note to the next in our musical system—isn't at all arbitrary. Each note in our musical system is equally spaced to our ears (but not necessarily to the ears of other species). Although there is not an equal change in cycles per second (Hz) as we climb from one note to the next, the distance between each note and the next sounds equal. How can this be? The frequency of each note in our system is approximately 6 percent more than the one before it. Our auditory system is sensitive both to relative changes and to proportional changes in sound. Thus, each increase in frequency of 6 percent gives us the impression that we have increased pitch by the same amount as we did last time.

The idea of proportional change is intuitive if you think about weights. If you're at a gym and you want to increase your weight lifting of the barbells from 5 pounds to 50 pounds, adding 5 pounds each week is not going to change the amount of weight you're lifting in an equal way. After a week of lifting 5 pounds, when you move to 10 you are doubling the weight; the next week when you move to 15 you are adding 1.5 times as much weight as you had before. An equal spacing—to give your muscles a similar increase of weight each week—would be to add a constant percentage of the previous week's weight each time you increase. For example, you might decide to add 50 percent each week, and so you would then go from 5 pounds to 7.5, then to 11.25, then to 16.83, and so on. The auditory system works the same way, and that is why our scale is based on a proportion: Every tone is 6 percent higher than the previous one, and when we increase each step by 6 percent twelve times, we end up having doubled our original frequency (the actual proportion is the twelfth root of two=1.059463...).

The twelve notes in our musical system are called the chromatic scale. Any scale is simply a set of musical pitches that have been chosen to be distinguishable from each other and to be used as the basis for constructing melodies.

In Western music we rarely use all the notes of chromatic scale in composition; instead, we use a subset of seven (or less often, five) of those twelve tones. Each of these subsets is itself a scale, and the type of scale we use has a large impact on the overall sound of a melody, and its emotional

sounding sad, or reflective, or even exotic. The most basic rock and country music songs use only major chords: “Johnny B. Goode,” “Blowin’ in the Wind,” “Honky Tonk Women,” and “Mamas Don’t Let Your Babies Grow Up to Be Cowboys,” for example.

Minor chords add complexity: in “Light My Fire” by the Doors, the verses are played in minor chords (“You know that it would be untrue . . .”) and then the chorus is played in major chords (“Come on baby, light my fire”). In “Jolene,” Dolly Parton mixes minor and major chords to give a melancholy sound. Steely Dan’s “Do It Again” (from the album *Can’t Buy a Thrill*) uses only minor chords.

Like single notes in the scale, chords also fall along a hierarchy of stability, depending on context. Certain chord progressions are part of every musical tradition, and even by the age of five, most children have internalized rules about what chord progressions are legal, or typical of their culture’s music; they can readily detect deviations from the standard sequences just as easily as we can detect when an English sentence is malformed, such as this one: “The pizza was too hot to sleep.” For brains to accomplish this, networks of neurons must form abstract representations of musical structure, and musical rules, something that they do automatically and without our conscious awareness. Our brains are maximally receptive—almost spongelike—when we’re young, hungrily soaking up any and all sounds they can and incorporating them into the very structure of our neural wiring. As we age, these neural circuits are somewhat less pliable, and so it becomes more difficult to incorporate, at a deep neural level, new musical systems, or even new linguistic systems.

* * *

Now the story about pitch becomes a bit more complicated, and it’s all the fault of physics. But this complication gives rise to the rich spectrum of sounds we hear in different instruments. All natural objects in the world have several modes of vibration. A piano string actually vibrates at several different rates at once. The same thing is true of bells that we hit with a hammer, drums that we hit with our hands, or flutes that we blow air into: The air molecules vibrate at several rates simultaneously, not just a single rate.

An analogy is the several types of motion of the earth that are simultaneously occurring. We know that the earth spins on its axis once every twenty-four hours, that it travels around the sun once every 365.25 days, and that the entire solar system is spinning along with the Milky Way galaxy. Several types of motion, all occurring at once. Another analogy is the many kinds of vibration that we often feel when riding a train. Imagine that you’re sitting on a train in an outdoor station, with the engine off. It’s windy, and you feel the car rock back and forth just a little bit. It does so with a regularity that you can time with your handy stopwatch, and you feel the train moving back and forth about twice a second. Next, the engineer starts the engine, and you feel a different kind of

vibration through your seat (due to the oscillations of the motor—pistons and crankshafts turning around at a certain speed). When the train starts moving, you experience a third sensation, the bump the wheels make every time they go over a track joint. Altogether, you will feel several different kinds of vibrations, all of them likely to be at different rates, or frequencies. When the train is moving, you are no doubt aware that there is vibration. But it is very difficult, if not impossible, for you to determine how many vibrations there are and what their rates are. Using specialized measuring instruments, however, one might be able to figure this out.

When a sound is generated on a piano, flute, or any other instrument—including percussion instruments like drums and cowbells—it produces many modes of vibration occurring simultaneously. When you listen to a single note played on an instrument, you’re actually hearing many, many pitches at once, not a single pitch. Most of us are not aware of this consciously, although some people can train themselves to hear this. The one with the slowest vibration rate—the one lowest in pitch—is referred to as the fundamental frequency, and the others are collectively called overtones.

To recap, it is a property of objects in the world that they generally vibrate at several different frequencies at once. Surprisingly, these other frequencies are often mathematically related to each other in a very simple way: as integer multiples of one another. So if you pluck a string and its slowest vibration frequency is one hundred times per second, the other vibration frequencies will be 2×100 (200 Hz), 3×100 Hz (300 Hz), etc. If you blow into a flute or recorder and cause vibrations at 310 Hz, additional vibrations will be occurring at twice, three times, four times, etc., this rate: 620 Hz, 930 Hz, 1240 Hz, etc. When an instrument creates energy at frequencies that are integer multiples such as this, we say that the sound is harmonic, and we refer to the pattern of energy at different frequencies as the overtone series. There is evidence that the brain responds to such harmonic sounds with synchronous neural firings—the neurons in auditory cortex responding to each of the components of the sound synchronize their firing rates with one another, creating a neural basis for the coherence of these sounds.

The brain is so attuned to the overtone series that if we encounter a sound that has all of the components except the fundamental, the brain fills it in for us in a phenomenon called *restoration of the missing fundamental*. A sound composed of energy at 100 Hz, 200 Hz, 300 Hz, 400 Hz, and 500 Hz is perceived as having a pitch of 100 Hz, its fundamental frequency. But if we artificially create a sound with energy at 200 Hz, 300 Hz, 400 Hz, and 500 Hz (leaving off the fundamental), we still perceive it as having a pitch of 100 Hz. We don’t perceive it as having a pitch of 200 Hz, because our brain “knows” that a normal, harmonic sound with a pitch of 200 Hz would have an overtone series of 200 Hz, 400 Hz, 600 Hz, 800 Hz, etc. We can also fool the brain by playing sequences that deviate from the overtone series such as this: 100 Hz, 210

Hz, 302 Hz, 405 Hz, etc. In cases like these, the perceived pitch shifts away from 100 Hz in a compromise between what is presented and what a normal harmonic series would imply.

When I was in graduate school, my advisor, Mike Posner, told me about the work of a graduate student in biology, Petr Janata. Although he hadn’t been raised in San Francisco like me, Petr had long bushy hair that he wore in a ponytail, played jazz and rock piano, and dressed in tie-dye: a true kindred spirit. Peter placed electrodes in the inferior colliculus of the barn owl, part of its auditory system. Then, he played the owls a version of Strauss’s “The Blue Danube Waltz,” made up of tones from which the fundamental frequency had been removed. Petr hypothesized that if the missing fundamental is restored at early levels of auditory processing, neurons in the owl’s inferior colliculus should fire at the rate of the missing fundamental. This was exactly what he found. And because the electrodes put out a small electrical signal with each firing—and because the firing rate is the same as a frequency of firing (as we saw above)—Petr sent the output of these electrodes to a small amplifier, and played back the sound of the owl’s neurons through a loudspeaker. What he heard was astonishing; the melody of “The Blue Danube Waltz” sang clearly from the loudspeakers: ba da da da, deet deet, deet deet. We were *hearing* the firing rates of the neurons and they were identical to the frequency of the missing fundamental. The overtone series had an instantiation not just in the early levels of auditory processing, but in a completely different species.

One could imagine an alien species that does not have ears, or that doesn’t have the same internal experience of hearing that we do. But it would be difficult to imagine an advanced species that had no ability whatsoever to sense vibrating objects. Where there is atmosphere there are molecules that vibrate in response to movement. And knowing whether something is generating noise or moving toward us or away from us, even when we can’t see it (because it is dark, our eyes aren’t attending to it, or we’re asleep) has a great survival value.

Because most physical objects cause molecules to vibrate in several modes at once, and because for many, many objects the modes bear simple integer relations to one another, the overtone series is a fact-of-the-world that we expect to find everywhere we look: in North America, in Fiji, on Mars, and on the planets orbiting Antares. Any organism that evolved in a world with vibrating objects is likely—given enough evolutionary time—to have evolved a processing unit in the brain that incorporated these regularities of its world. Because pitch is a fundamental cue to an object’s identity, we would expect to find tonotopic mappings as we do in human auditory cortex, and synchronous neural firings for tones that bear octave and other harmonic relations to one another; this would help the brain (alien or terrestrial) to figure out that all these tones probably originated from the same object.

The overtones are often referred to by numbers: The first overtone is

the first vibration frequency above the fundamental, the second overtone is the second vibration frequency above the fundamental, etc. Because physicists like to make the world confusing for the rest of us, there is a parallel system of terminology called harmonics, and I think it was designed to make undergraduates go crazy. In the lingo of harmonics, the first harmonic is the fundamental frequency, the second harmonic is equal to the first overtone, and so on. Not all instruments vibrate in modes that are so neatly defined. Sometimes, as with the piano (because it is a percussive instrument), the overtones can be close, but not exact, multiples of the fundamental frequency, and this contributes to their characteristic sound. Percussion instruments, chimes, and other objects—depending on composition and shape—often have overtones that are clearly not integer multiples of the fundamental, and these are called partials or inharmonic overtones. Generally, instruments with inharmonic overtones lack the clear sense of pitch that we associate with harmonic instruments, and the cortical basis for this may relate to a lack of synchronous neural firing. But they still do have a sense of pitch, and we hear this most clearly when we can play inharmonic notes in succession. Although you may not be able to hum along with the sound of a single note played on a woodblock or a chime, we can play a recognizable melody on a set of woodblocks or chimes because our brain focuses on the changes in the overtones from one to another. This is essentially what is happening when we hear people playing a song on their cheeks.

A flute, a violin, a trumpet, and a piano can all play the same tone—that is, you can write a note on a musical score and each instrument will play a tone with an identical fundamental frequency, and we will (tend to) hear an identical pitch. But these instruments all sound very different from one another.

This difference is timbre, and it is the most important and ecologically relevant feature of auditory events. The timbre of a sound is the principal feature that distinguishes the growl of a lion from the purr of a cat, the crack of thunder from the crash of ocean waves, the voice of a friend from that of a bill collector one is trying to dodge. Timbral discrimination is so acute in humans that most of us can recognize hundreds of different voices. We can even tell whether someone close to us—our mother, our spouse—is happy or sad, healthy or coming down with a cold, based on the timbre of that voice.

Timbre is a consequence of the overtones. Different materials have different densities. A piece of metal will tend to sink to the bottom of a pond; an identically sized and shaped piece of wood will float. Partly due to density, and partly due to size and shape, different objects also make different noises when you strike them with your hand, or gently tap them with a hammer. Imagine the sound that you'd hear if you tap a hammer (gently, please!) against a guitar—a hollow, wooden *plunk* sound. Or if you tap a piece of metal, like a saxophone—a tinny *plink*. When you tap these objects, the energy from the hammer causes the molecules within

them to vibrate, to dance at several different frequencies, frequencies determined by the material the object is made out of, its size, and its shape. If the object is vibrating at, say, 100 Hz, 200 Hz, 300 Hz, 400 Hz, etc., the intensity of vibration doesn't have to be the same for each of these harmonics, and in fact, typically, it is not.

When you hear a saxophone playing a tone with a fundamental frequency of 220 Hz, you are actually hearing many tones, not just one. The other tones you hear are integer multiples of the fundamental: 440, 660, 880, 1100, 1320, 1540, etc. These different tones—the overtones—have different intensities, and so we hear them as having different loudnesses. The particular pattern of loudnesses for these tones is distinctive of the saxophone, and they are what give rise to its unique tonal color, its unique sound—its timbre. A violin playing the same written note (220 Hz) will have overtones at the same frequencies, but the pattern of how loud each one is with respect to the others will be different. Indeed, for each instrument, there exists a unique pattern of overtones. For one instrument, the second overtone might be louder than in another, while the fifth overtone might be softer. Virtually all of the tonal variation we hear—the quality that gives a trumpet its trumpetiness and that gives a piano its pianoness—comes from the unique way in which the loudnesses of the overtones are distributed.

Each instrument has its own overtone profile, which is like a fingerprint. It is a complicated pattern that we can use to identify the instrument. Clarinets, for example, are characterized by having relatively high amounts of energy in the odd harmonics—three times, five times, and seven times the multiples of the fundamental frequency, etc. (This is a consequence of their being a tube that is closed at one end and open at the other.) Trumpets are characterized by having relatively even amounts of energy in both the odd and the even harmonics (like the clarinet, the trumpet is also closed at one end and open at the other, but the mouthpiece and bell are designed to smooth out the harmonic series). A violin that is bowed in the center will yield mostly odd harmonics and accordingly can sound similar to a clarinet. But bowing one third of the way down the instrument emphasizes the third harmonic and its multiples: the sixth, the ninth, the twelfth, etc.

All trumpets have a timbral fingerprint, and it is readily distinguishable from the timbral fingerprint for a violin, piano, or even the human voice. To the trained ear, and to most musicians, there even exist differences among trumpets—all trumpets don't sound alike, nor do all pianos or all accordions. What distinguishes one particular piano from another is that their overtone profiles will differ slightly from each other, but not, of course, as much as they will differ from the profile for a harpsichord, organ, or tuba. Master musicians can hear the difference between a Stradivarius violin and a Guarneri within one or two notes. I can hear the difference between my 1936 Martin 000-18 acoustic guitar, my 1973 Martin D-18, and my 1996 Collings D2H very clearly; they sound

like different instruments, even though they are all acoustic guitars; I would never confuse one with another. That is timbre.

Natural instruments—that is, acoustic instruments made out of real-world materials such as metal and wood—tend to produce energy at several frequencies at once because of the way the internal structure of their molecules vibrates. Suppose that I invent an instrument that, unlike any natural instruments we know of, produces energy at one, and only one, frequency. Let's call this hypothetical instrument a generator (because it can generate tones of specific frequencies). If I line up a bunch of generators, I could set each one of them to play a specific frequency corresponding to the overtone series for a particular instrument playing a particular tone. I could have a bank of these generators making sounds at 110, 220, 330, 440, 550, and 660 Hz, which would give the listener the impression of a 110 Hz tone played by a musical instrument. Furthermore, I could control the amplitude of each of my generators and make each of the tones play at a particular loudness, corresponding to the overtone profile of a natural musical instrument. If I did that, the resulting bank of generators would approximate the sound of a clarinet, or flute, or any other instrument I was trying to emulate.

Additive synthesis such as the above approach achieves a synthetic version of a musical-instrument timbre by adding together elemental sonic components of the sound. Many pipe organs, such as those found in churches, have a feature that will let you play around with this. On most pipe organs you press a key (or a pedal), which sends a blast of air through a metal pipe. The organ is constructed of hundreds of pipes of different sizes, and each one produces a different pitch, corresponding to its size, when air is shot through it; you can think of them as mechanical flutes, in which the air is supplied by an electric motor rather than by a person blowing. The sound that we associate with a church organ—its particular timbre—is a function of there being energy at several different frequencies at once, just as with other instruments. Each pipe of the organ produces an overtone series, and when you press a key on the organ keyboard, a column of air is blasted through more than one pipe at a time, giving a very rich spectrum of sounds. These supplementary pipes, in addition to the one that vibrates at the fundamental frequency of the tone you're trying to play, either produce tones that are integer multiples of the fundamental frequency, or are closely related to it mathematically and harmonically.

The organ player typically has control over which of these supplementary pipes he wants to blow air through by pulling and pushing levers, or drawbars, that direct the flow of air. Knowing that clarinets have a lot of energy in the odd harmonics of the overtone series, a clever organ player could simulate the sound of a clarinet by manipulating drawbars in such a way as to re-create the overtone series of that instrument. A little bit of 220 Hz here, a dash of 330 Hz, a dollop of 440 Hz, a heaping helping of 550 Hz, and *voilà!*—you've cooked yourself up a

reasonable facsimile of an instrument.

Starting in the late 1950s, scientists began experimenting with building such synthesis capabilities into smaller, more compact electronic devices, creating a family of new musical instruments known collectively as synthesizers. By the 1960s, synthesizers could be heard on records by the Beatles (on "Here Comes the Sun" and "Maxwell's Silver Hammer") and Walter/Wendy Carlos (*Switched-On Bach*), followed by groups who sculpted their sound around the synthesizer, such as Pink Floyd and Emerson, Lake and Palmer.

Many of these synthesizers used additive synthesis as I've described it here, and later ones used more complex algorithms such as wave guide synthesis (invented by Julius Smith at Stanford) and FM synthesis (invented by John Chowning at Stanford). But merely copying the overtone profile, while it can create a sound reminiscent of the actual instrument, yields a rather pale copy. There is more to timbre than just the overtone series. Researchers still argue about what this "more" is, but it is generally accepted that, in addition to the overtone profile, timbre is defined by two other attributes that give rise to a perceptual difference from one instrument to another: attack and flux.

Stanford University sits on a bucolic stretch of land just south of San Francisco and east of the Pacific Ocean. Rolling hills covered with pastureland lie to the west, and the fertile Central Valley of California is just an hour or so to the east, home of a large proportion of the world's raisins, cotton, oranges, and almonds. To the south, near the town of Gilroy, are vast fields of garlic. Also to the south is Castroville, known as the "artichoke capital of the world." (I once suggested to the Castroville Chamber of Commerce that they change *capital* to *heart*. The response was not enthusiastic.)

Stanford has become something of a second home for computer scientists and engineers who love music. John Chowning, who was well known as an avant-garde composer, has had a professorship in the music department there since the 1970s, and was among a group of pioneering composers at the time who were using the computer to create, store, and reproduce sounds in their compositions. Chowning later became the founding director of the Center for Computer Research in Music and Acoustics at Stanford, known as CCRMA (pronounced CAR-ma; insiders joke that the first c is silent). Chowning is warm and friendly. When I was an undergraduate at Stanford, he would put his hand on my shoulder and ask what I was working on. You got the feeling talking to a student was for him an opportunity to learn something. In the early 1970s, while fiddling with the computer and with sine waves—the sorts of artificial sounds that are made by computers and used as the building blocks of additive synthesis—Chowning noticed that changing the frequency of these waves as they were playing created sounds that were musical. By controlling these parameters just so, he was able to simulate the sounds of a number of musical instruments. This new technique became known

as frequency modulation synthesis, or FM synthesis, and became embedded first in the Yamaha DX9 and DX7 line of synthesizers, which revolutionized the music industry from the moment of their introduction in 1983. FM synthesis democratized music synthesis. Before FM, synthesizers were expensive, clunky, and hard to control. Creating new sounds took a great deal of time, experimentation, and know-how. But with FM, any musician could obtain a convincing instrumental sound at the touch of a button. Songwriters and composers who could not afford to hire a horn section or an orchestra could now play around with these textures and sounds. Composers and orchestrators could test out arrangements before taking the time of an entire orchestra to see what worked and what didn't. New Wave bands like the Cars and the Pretenders, as well as mainstream artists like Stevie Wonder, Hall and Oates, and Phil Collins, started to use FM synthesis widely in their recordings. A lot of what we think of as "the eighties sound" in popular music owes its distinctiveness to the particular sound of FM synthesis.

With the popularization of FM came a steady stream of royalty income that allowed Chowning to build up CCRMA, attracting graduate students and top-flight faculty members. Among the first of many famous electronic music/music-psychology celebrities to come to CCRMA were John R. Pierce and Max Mathews. Pierce had been the vice president of research at the Bell Telephone Laboratories in New Jersey, and supervised the team of engineers who built and patented the transistor—and it was Pierce who named the new device (TRANSISTOR resISTOR). In his distinguished career, he also is credited with inventing the traveling wave vacuum tube, and launching the first telecommunications satellite, Telstar. He was also a respected science fiction writer under the pseudonym J. J. Coupling. Pierce created a rare environment in any industry or research lab, one in which the scientists felt empowered to do their best and in which creativity was highly valued. At the time, the Bell Telephone Company/AT&T had a complete monopoly on telephone service in the U.S. and a large cash reserve. Their laboratory was something of a playground for the very best and brightest inventors, engineers, and scientists in America. In the Bell Labs "sandbox," Pierce allowed his people to be creative without worrying about the bottom line or the applicability of their ideas to commerce. Pierce understood that the only way true innovation can occur is when people don't have to censor themselves and can let their ideas run free. Although only a small proportion of those ideas may be practical, and a smaller proportion still would become products, those that did would be innovative, unique, and potentially very profitable. Out of this environment came a number of innovations including lasers, digital computers, and the Unix operating system.

I first met Pierce in 1990 when he was already eighty and was giving lectures on psychoacoustics at CCRMA. Several years later, after I had earned my Ph.D. and moved back to Stanford, we became friends and

would go out to dinner every Wednesday night and discuss research. He once asked me to explain rock and roll music to him, something he had never paid any attention to and didn't understand. He knew about my previous career in the music business, and he asked if I could come over for dinner one night and play six songs that captured all that was important to know about rock and roll. Six songs to capture all of rock and roll? I wasn't sure I could come up with six songs to capture the Beatles, let alone all of rock and roll. The night before he called to tell me that he had heard Elvis Presley, so I didn't need to cover that.

Here's what I brought to dinner:

- 1) "Long Tall Sally," Little Richard
- 2) "Roll Over Beethoven," the Beatles
- 3) "All Along the Watchtower," Jimi Hendrix
- 4) "Wonderful Tonight," Eric Clapton
- 5) "Little Red Corvette," Prince
- 6) "Anarchy in the U.K.," the Sex Pistols

A couple of the choices combined great songwriters with different performers. All are great songs, but even now I'd like to make some adjustments. Pierce listened and kept asking who these people were, what instruments he was hearing, and how they came to sound the way they did. Mostly, he said, he liked the timbres of the music. The songs themselves and the rhythms didn't interest him that much, but he found the timbres to be remarkable—new, unfamiliar, and exciting. The fluid romanticism of Clapton's guitar solo in "Wonderful Tonight," combined with the soft, pillowy drums. The sheer power and density of the Sex Pistols' brick-wall-of-guitars-and-bass-and-drums. The sound of a distorted electric guitar wasn't all that was new to Pierce. The ways in which instruments were combined to create a unified whole—bass, drums, electric and acoustic guitars, and voice—that was something he had never heard before. Timbre was what defined rock for Pierce. And it was a revelation to both of us.

The pitches that we use in music—the scales—have remained essentially unchanged since the time of the Greeks, with the exception of the development—really a refinement—of the equal tempered scale during the time of Bach. Rock and roll may be the final step in a millennium-long musical revolution that gave perfect fourths and fifths a prominence in music that had historically been given only to the octave. During this time, Western music was largely dominated by pitch. For the past two hundred years or so, timbre has become increasingly important. A standard component of music across all genres is to restate a melody using different instruments—from Beethoven's Fifth and Ravel's *Bolero* to the Beatles' "Michelle" and George Strait's "All My Ex's Live in Texas." New musical instruments have been invented so that composers might have a larger palette of timbral colors from which to draw. When a country or popular singer stops singing and another instrument takes up

the melody—even without changing it in any way—we find pleasurable the repetition of the same melody with a different timbre.

* * *

The avant-garde composer Pierre Schaeffer (pronounced Sheh-FEHR, using your best imitation of a French accent) performed some crucial experiments in the 1950s that demonstrated an important attribute of timbre in his famous “cut bell” experiments. Schaeffer recorded a number of orchestral instruments on tape. Then, using a razor blade, he cut the beginnings off of these sounds. This very first part of a musical instrument sound is called the attack; this is the sound of the initial hit, strum, bowing, or blowing that causes the instrument to make sound.

The gesture our body makes in order to create sound from an instrument has an important influence on the sound the instrument makes. But most of that dies away after the first few seconds. Nearly all of the gestures we make to produce a sound are impulsive—they involve short, punctuated bursts of activity. In percussion instruments, the musician typically does not remain in contact with the instrument after this initial burst. In wind instruments and bowed instruments, on the other hand, the musician continues to be in contact with the instrument after the initial impulsive contact—the moment when the air burst first leaves her mouth or the bow first contacts the string; the continued blowing and bowing has a smooth, continuous, and less impulsive quality.

The introduction of energy to an instrument—the attack phase—usually creates energy at many different frequencies that are not related to one another by simple integer multiples. In other words, for the brief period after we strike, blow into, pluck, or otherwise cause an instrument to start making sound, the impact itself has a rather noisy quality that is not especially musical—more like the sound of a hammer hitting a piece of wood, say, than like a hammer hitting a bell or a piano string, or like the sound of wind rushing through a tube. Following the attack is a more stable phase in which the musical tone takes on the orderly pattern of overtone frequencies as the metal or wood (or other material) that the instrument is made out of starts to resonate. This middle part of a musical tone is referred to as the steady state—in most instances the overtone profile is relatively stable while the sound emanates from the instrument during this time.

After Schaeffer edited out the attack of orchestral instrument recordings, he played back the tape and found that it was nearly impossible for most people to identify the instrument that was playing. Without the attack, pianos and bells sounded remarkably unlike pianos and bells, and remarkably similar to one another. If you splice the attack of one instrument onto the steady state, or body, from another, you get varied results: In some cases, you hear an ambiguous hybrid instrument that sounds more like the instrument that the attack came from than the

one the steady state came from. Michelle Castellengo and others have discovered that you can create entirely new instruments this way; for example, splicing a violin bow sound onto a flute tone creates a sound that strongly resembles a hurdy-gurdy street organ. These experiments showed the importance of the attack.

The third dimension of timbre—flux—refers to how the sound changes after it has started playing. A cymbal or gong has a lot of flux—its sound changes dramatically over the time course of its sound—while a trumpet has less flux—its tone is more stable as it evolves. Also, instruments don’t sound the same across their range. That is, the timbre of an instrument sounds different when playing high and low notes. When Sting reaches up toward the top of his vocal range in “Roxanne” (by The Police), his straining, reedy voice conveys a type of emotion that he can’t achieve in the lower parts of his register, such as we hear on the opening verse of “Every Breath You Take,” a more deliberate, longing sound. The high part of Sting’s register pleads with us urgently as his vocal cords strain, the low part suggests a dull aching that we feel has been going on for a long time, but has not yet reached the breaking point.

Timbre is more than the different sounds that instruments make. Composers use timbre as a compositional tool; they choose musical instruments—and combinations of musical instruments—to express particular emotions, and to convey a sense of atmosphere or mood. There is the almost comical timbre of the bassoon in Tchaikovsky’s *Nutcracker Suite* as it opens the “Chinese Dance,” and the sensuousness of Stan Getz’s saxophone on “Here’s That Rainy Day.” Substitute a piano for the electric guitars in the Rolling Stones’ “Satisfaction” and you’d have an entirely different animal. Ravel used timbre as a compositional device in *Bolero*, repeating the main theme over and over again with different timbres; he did this after he suffered brain damage that impaired his ability to hear pitch. When we think of Jimi Hendrix, it is the timbre of his electric guitars and his voice that we are likely to recall the most vividly.

Composers such as Scriabin and Ravel talk about their works as sound paintings, in which the notes and melodies are the equivalent of shape and form, and the timbre is equivalent to the use of color and shading. Several popular songwriters—Stevie Wonder, Paul Simon, and Lindsey Buckingham—have described their compositions as sound paintings, with timbre playing a role equivalent to the one that color does in visual art, separating melodic shapes from one another. But one of the things that makes music different from painting is that it is dynamic, changing across time, and what moves the music forward are rhythm and meter. Rhythm and meter are the engine driving virtually all music, and it is likely that they were the very first elements used by our ancestors to make protomusics, a tradition we still hear today in tribal drumming, and in the rituals of various preindustrial cultures. While I believe timbre is now at the center of our appreciation of music, rhythm has held supreme power over listeners for much longer.