BOX 4.2
ERPs reveal statistical skills in newborns

The head-turn preference paradigm (see Method 4.1) is a clever behavioral method that has allowed researchers to test infants' knowledge without requiring any sophisticated responses or behaviors. Nevertheless, it does require babies to have developed the neck muscles that are needed to turn their heads in response to a stimulus. It also requires the babies to sustain full consciousness for reasonable periods of time. This makes it challenging to study the learning skills of newborn babies, with their floppy necks and tendency to sleep much of the time when they're not actively feeding. But as you saw in Chapter 3, ERPs (event-related potentials) can be used to probe the cognitive processes of people in a vegetative state, bypassing the need for any meaningful behavior at all in response to stimuli. Could the same method be used to assess the secret cognitive life of newborns?

Tuomas Teinonen and colleagues (2009) used ERP methods to test whether newborns can pick up on the transitional probabilities of syllables in a sample of speech. Their subjects were less than 2 days old, and they listened to at least 15 minutes of running speech consisting of ten different three-syllable made-up words randomly strung together. After this 15-minute "learning" period, the researchers analyzed the electrical activity in the babies' brains. Because the ERPs of newborn babies are less wildly variable if measured during sleep, the researchers limited the analysis to brain activity that was monitored during active sleep—which turned out to represent 40%–80% of the hour-long experiment.

ERP activity was compared for each of the three syllables of the novel "words." The logic behind this comparison was that, since the first syllable for any given "word" was less predictable (having a lower transitional probability) than the second and third syllables, it should show heightened brain activity compared with the other two syllables. **Figure 4.4** shows the results of the study, which indicate a region of enhanced negative activity for the first of the three syllables. Similar results have since shown that newborns can also track the transitional probabilities of tones (Kudo et al., 2011). These remarkable studies reveal that the ability to pull statistical regularities from the auditory world is a robust skill that's available to humans from the very first moments after birth.

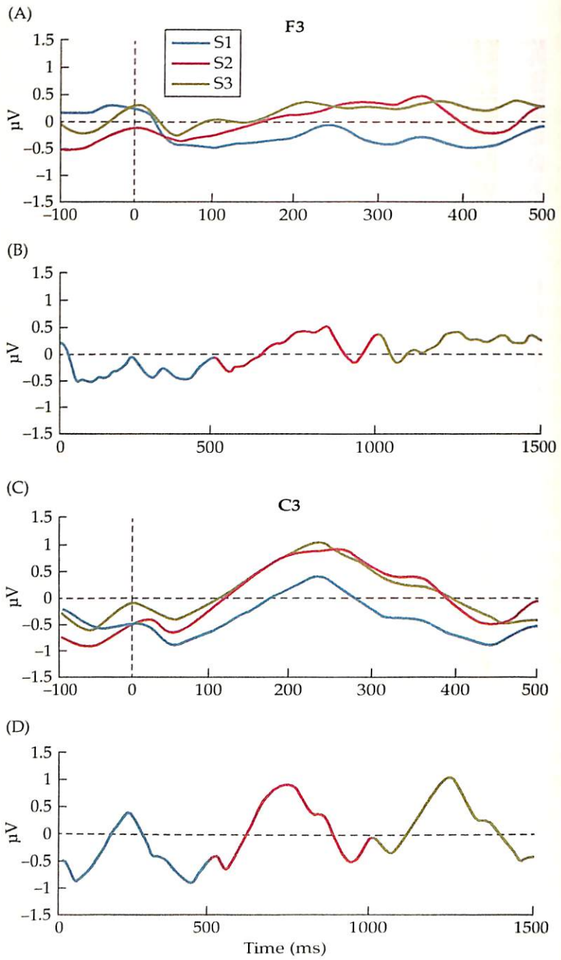


Figure 4.4 ERP activity at two recording sites (F3 and C3) shows enhanced negativity. In panels (A) and (C), the syllables are aligned so that each syllable's onset corresponds to 0. The shaded areas show the region where there is a statistically significant difference between the first syllable (S1) and the second and third syllables (S2 and S3). Panels (B) and (D) track EEG activity for the three syllables spoken in sequence. (Adapted from Teinonen et al., 2009.)

regularities within stimuli as diverse as musical tones (Saffran et al., 1999) and visual shapes (Fiser & Aslin, 2001). It appears, then, that one of the earliest language-related tasks that a baby undertakes rests on a pretty sturdy and highly general cognitive skill that we share with animals as we all try to make sense of the world around us. Indeed, it's likely that as humans, we literally have this ability from birth (see Box 4.2).

But that's not the end of the story. Just because individuals of different species can track statistical regularities across a number of different domains doesn't necessarily mean that the same *kinds* of regularities are being tracked in all these cases. In fact, Toro and Trobalón found that rats were able to use simple statistical cues to segment speech but weren't sensitive to some of the more complex cues that have been found to be used by human infants. And there may also be some subtle distinctions in the kinds of cues that are used in dealing with language, for example, as opposed to other, non-linguistic stimuli. These more nuanced questions are taken up in Digging Deeper at the end of this chapter.

4.3 What Are the Sounds?

How many distinct sounds are there in a language?

You might think that having to figure out where the words are in your language is hard enough. But in fact, if we back up even more, it becomes apparent that babies are born without even knowing what *sounds* make up their language. These sounds, too, have to be learned. This is not as trivial as it seems. As an adult whose knowledge of your language is deeply entrenched, you have the illusion that there's a fairly small number of sounds that English speakers produce (say, about 40), and that it's just a matter of learning what these 40 or so sounds are. But in truth, English speakers produce many more than 40 sounds.

Here's an example: Ever since your earliest days in school, when you were likely given exercises to identify sounds and their corresponding letter symbols, you learned that the words *tall* and *tree* begin with the same sound, and that the second and third consonants of the word *potato* are identical. But that's not exactly right. Pay close attention to what's happening with your tongue as you say these sounds the way you normally would in conversational speech, and you'll see that not all consonants that are represented by the letter *t* are identical. For example, you likely said *tree* using a sound that's a lot like the first sound in *church*, and unless you were fastidiously enunciating the word potato, the two "t" sounds were not the same. It turns out that sounds are affected by the phonetic company they keep. And these subtle distinctions matter. If you were to cut out the "t" in *tall* and swap it for the "t" in *tree*, you would be able to tell the difference. The resulting word would sound a bit weird. The sound represented by the symbol *t* also varies depending on whether it's placed at the very beginning of a syllable, as in *tan*, or is the second member of a consonant cluster, as in *Stan*.

Not convinced? Here's some playing with fire you're encouraged to try at home. Place a lit match a couple of inches away from your mouth and say the word *Stan*. Now say *tan*. If the match is at the right distance from your mouth (and you might need to play around with this a bit), it will be puffed out when you say *tan*, but not when you say *Stan*. When you use "t" at the beginning of a syllable, you release an extra flurry of air. You can feel this if you hold your palm up close to your mouth while saying these words.

These kinds of variations are in no way limited to the "t" sound in English; any and all of the 40-odd sounds of English can be and are produced in a variety of different ways, depending on which sounds they're keeping company

WEB ACTIVITY 4.4

Scrambled speech In this demo, you'll get a sense of what speech is like when different versions of sounds that we normally think of as the same have been scrambled from their normal locations in words.

with. Suddenly, the inventory of approximately 40 sounds has mushroomed into many more.

Not only do the surrounding sounds make a difference to how any given sound is pronounced, but so do things like how fast the speaker is talking; whether the speaker is male or female, old or young; whether the speaker is shouting, whispering, or talking at a moderate volume; and whether he or she is talking to a baby or a friend at a bar, or is reading the news on a national television network.

And yet, despite all this variation, we do have the sense that all "t" sounds, regardless of how they're made, should be classified as representing *one* kind of sound. This sense goes beyond just knowing that all of these "t" instances are captured by the orthographic symbol *t* or *T*. More to the point, while swapping out one kind of "t" sound for another might sound weird, it doesn't change what *word* has been spoken. Not like swapping out the "t" in *ten* for a "d" sound, for example. Now, all of a sudden, you have a completely different word, *den*, with a completely different meaning. This means that not all sound distinctions are created equal. Some change the fundamental identity of a speech sound, while others are the speech equivalent of sounds putting on different outfits depending on which other sounds they're hanging out with, or what event they happen to be at.

When a sound distinction has the potential to actually cause a change in meaning, that distinction yields separate **phonemes**. But when sound differences don't fundamentally change the identity of a speech unit, we say they create different **allophones** of the same phoneme. You know that sound distinctions create different phonemes when it's possible to create **minimal pairs** of words in which changing a single sound results in a change in meaning. For example, the difference between "t" and "d" is a phonemic distinction, not an allophonic distinction, because we get minimal pairs such as *ten, den*; *toe, doe*; and *bat, bad* (see Table 4.1).

Our impression is that the difference between the sounds "t" and "d" is a big one, while the difference between the two "t" sounds in *tan* and *Stan* is very slight and hard to hear. But this sense is merely a product of the way we mentally categorize these sounds. *Objectively*, the difference between the sounds in both pairs is close to exactly the same, and as you'll see later on, there's evidence that we're not deaf to the acoustic differences between allophones—but we've mentally amplified the differences between phonemes, and minimized the differences between allophones.

phoneme The smallest unit of sound that changes the meaning of a word; often identified by forward slashes; e.g., /t/ is a phoneme because replacing it in the word *tan* makes a different word .

allophones Two or more similar sounds that are variants of the same phoneme; often identified by brackets; e.g., [t] and [tʰ] represent the two allophones of /t/ in the words *Stan* and *tan*.

minimal pair A pair of words that have different meanings, but all of the same sounds with the exception of one phoneme; e.g., *tan* and *man*.

A catalogue of sound distinctions

To begin to describe differences among speech sounds in a more objective way, it's useful to break them down into their characteristics. This turns out to be quite easy to do, because speech sounds vary systematically along a fairly small number of dimensions. For example, we only need three dimensions to capture most consonants of English and other languages: place of articulation, manner of articulation, and voicing.

TABLE 4.1 Examples of minimal word pairs[a]

pad/**b**ad	sa**f**e/sa**v**e	bi**gg**er/bi**dd**er	**m**eet/**n**eat	bu**sh**/bu**tch**	go**n**e/go**ng**
ta**p**/ta**b**	le**t**/le**d**	**c**all/**t**all	me**m**e/mea**n**	**ch**in/**g**in	**y**ell/**w**ell
fan/**v**an	la**g**/la**d**	ba**k**e/bai**t**	**sh**in/**ch**in	**r**ead/**l**ead	**w**ell/**h**ell

[a]In English, the presence of minimal word pairs that differ only with respect to a single sound shows that those sounds (boldface type) are distinct phonemes. Be sure to focus on how the words sound rather than on how they are spelled.

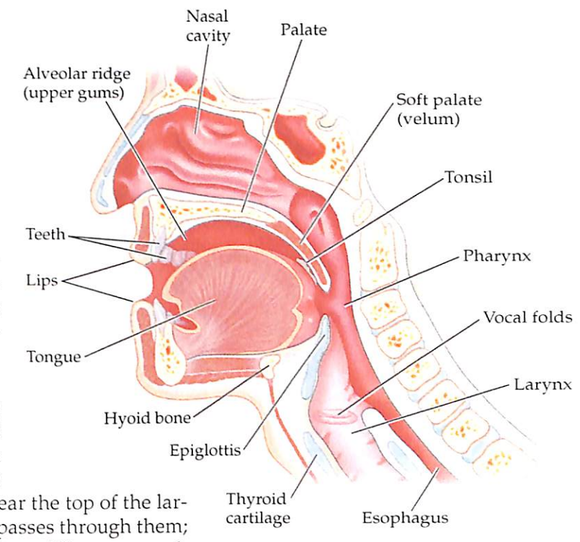


Figure 4.5 The human vocal tract, showing the various articulators. Air from the lungs passes through the larynx and over the vocal folds, making the folds vibrate and thus producing sound waves. The tongue, lips, and teeth help form this sound into speech. The place of articulation refers to the point at which the airflow becomes obstructed; for example, if airflow is briefly cut off by placing the tongue against the alveolar ridge, a sound would be said to be alveolar; a sound made by obstructing airflow at the velum would be velar.

PLACE OF ARTICULATION Consonants are typically made by pushing air out of the lungs, through the larynx and the **vocal folds** (often called the "vocal cords," although the term "folds" is much more accurate) and through the mouth or nose (see Figure 4.5). The vocal folds, located near the top of the larynx, are a pair of loosely attached flaps that vibrate as air passes through them; these vibrations produce sound waves that are shaped into different speech sounds by the rest of the vocal tract. (To hear your vocal folds in action, try first whispering the syllable "aahh," and then utter it as you normally would—the "noise" that's added to the fully sounded vowel comes from the vocal fold vibration, or **phonation**.) To create a consonant sound, the airflow passing through the vocal tract has to be blocked—either partially or completely—at some point above the larynx. The location where this blockage occurs has a big impact on what the consonant sounds like. For example, both the "p" and "t" sounds completely block the airflow for a short period of time. But the "p" sound is made by closing the air off at the lips, while "t" is made by closing it off at the little ridge just behind your teeth, or the alveolar ridge. A sound like "k," on the other hand, is made by closing the air off at the back of the mouth, touching the palate with the back of the tongue rather than its tip. And these are really the only significant differences between these sounds. (As you'll see in a moment, along the other two sound dimensions, "p," "t," and "k" are all alike.) Moving from lips to palate, the sounds are described as **bilabial** for "p," **alveolar** for "t," and **velar** for "k." Other intermediate places exist as well, as described next and summarized in Figure 4.6.

vocal folds Also known as "vocal cords," these are paired "flaps" in the larynx that vibrate as air passes over them. The vibrations are shaped into speech sounds by the other structures (tongue, alveolar ridge, velum, etc.) of the vocal tract.

phonation Production of sound by the vibrating vocal folds.

bilabial Describes a sound that is produced by obstructing airflow at the lips.

Place of articulation

Manner of articulation	Bilabial	Labio-dental	Inter-dental	Alveolar	Alveo-palatal	Palatal	Velar	Glottal
Stop	p b			t d			k g	ʔ
Fricative		f v	θ ð	s z	ʃ ʒ			h
Affricative					ʧ ʤ			
Nasal	m			n			ŋ	
Lateral liquid				l				
Retroflex liquid				ɹ				
Glide	ʍ w					j		

State of the glottis
Voiceless
Voiced

Figure 4.6 A chart of the consonant phonemes of Standard American English. In this presentation, the sounds are organized by place of articulation, manner of articulation, and voicing. (From the International Phonetic Association.)

MANNER OF ARTICULATION As mentioned, the airflow in the vocal tract can be obstructed either completely or partially. When the airflow is stopped completely somewhere in the mouth, you wind up producing what is known as a **stop consonant**. Stop consonants come in two varieties. If the air is fully blocked in the mouth and not allowed to leak out through the nose, you have an **oral stop**—our old friends "p," "t," and "k." But if you lower the velum (the soft tissue at the back of the roof of your mouth; see Figure 4.5) in a way that lets the air pass through your nose, you'll produce a **nasal stop**, which includes sounds like "m," "n," and the "ŋ" sound in words like *sing* or *fang*. You might have noticed that when your nose is plugged due to a cold, your nasal stops end up sounding like oral stops—"my nose" turns into "by dose" because no air can get out through your stuffed-up nose.

But your tongue is capable of more subtlety than simply blocking airflow entirely when some part of it is touched against the oral cavity. It can also *narrow* the airflow in a way that produces a turbulent sound—such as "s" or "f" or "z." These turbulent sounds are called **fricatives**. If you squish an oral stop and a fricative together, like the first and last consonants in *church* or *judge*, you wind up with an **affricate**.

Or, you can let air escape over both sides of your tongue, producing what are described as **liquid sounds** like "l" or "r," which differ from each other only in whether the blade (the front third) of your tongue is firmly planted against the roof of your mouth or is bunched back.

Finally, if you obstruct the airflow only mildly, allowing most of it to pass through the mouth, you will produce a **glide**. Pucker your lips, and you'll have a "w" sound, whereas if you place the back of your tongue up toward the velum as if about to utter a "k" but stop well before the tongue makes contact, you'll produce a "y" sound.

VOICING The last sound dimension has to do with whether (and when) the vocal folds are vibrating as you utter a consonant. People commonly refer to this part of the human anatomy as the "vocal cords" because, much like a musical instrument (such as a violin or cello) that has strings or cords, pitch in the human voice is determined by how quickly this vocal apparatus vibrates. But unlike a cello, voice isn't caused by passing something over a set of strings to make them vibrate. Rather, sound generation in the larynx (the "voice box") involves the "flaps" of the vocal folds (see Figure 4.5), which can constrict either loosely or very tightly. Sound is made when air coming up from the lungs passes through these flaps; depending on how constricted the vocal folds are, you get varying amounts of vibration, and hence higher or lower pitch. Think of voice as less like a cello and more like air flowing through the neck of a balloon held either tightly or loosely (though, in terms of beauty, I'll grant that the human voice is more like a cello than like a rapidly deflating balloon).

Vowels, unless whispered (or in certain special situations), are almost always produced while the vocal folds are vibrating. But consonants can vary. Some, like "z," "v," and "d," are made with vibrating vocal folds, while others, like "s," "f," and "t," are not—try putting your hand up against your throat just above your Adam's apple, and you'll be able to feel the difference.

Oral stops are especially interesting when it comes to voicing. Remember that for these sounds, the airflow is completely stopped somewhere in the mouth when two articulators come together—whether two lips, or a part of the tongue and the roof of the mouth. Voicing refers to when the vocal folds begin to vibrate relative to this closure and release. When vibration happens just about simultaneously with the release of the articulators (say, within about 20

alveolar Describes a sound whose place of articulation is the alveolar ridge, just behind the teeth.

velar Describes a sound whose place of articulation is the velum (the soft tissue at the back of the roof of your mouth; see Figure 4.5).

stop consonant A sound produced when airflow is stopped completely somewhere in the vocal tract.

oral stop A stop consonant made by fully blocking air in the mouth and not allowing it to leak out through the nose; e.g., "p," "t," and "k."

nasal stop A stop consonant made by lowering the velum in a way that lets the air pass through your nose; e.g., "m," "n," and the "ŋ" sound in words like *sing* or *fang*.

fricative A sound that is produced when your tongue narrows the airflow in a way that produces a turbulent sound; e.g., "s," "f," or "z."

affricate A sound that is produced when you combine an oral stop and a fricative together, like the first and last consonants in *church* or *judge*.

liquid sound A sound that is produced when you let air escape over both sides of your tongue; e.g., "l" or "r."

glide A sound that is produced when you obstruct the airflow only mildly, allowing most of it to pass through the mouth; e.g., "w" or "y."

LANGUAGE AT LARGE 4.1

The articulatory phonetics of beatboxing

It's not likely that a YouTube video of someone reciting a random list of words from the *Oxford English Dictionary* would spread virally—the act is just not that interesting. But many people *are* rightly riveted by the skills of virtuoso beatbox artists. Beatboxing is the art of mimicking musical and percussive sounds, and during their performances beatboxers routinely emit sounds with names like *808 snare drum roll*, *brushed cymbal*, *reverse classic kick drum*, *bongo drum*, and *electro scratch*. When you see them in action, what comes out of their mouths seems more machine-like than human.

And yet, when you look at how these sounds are actually made, it becomes clear that the repertoire of beatbox sounds is the end result of creatively using and recombining articulatory gestures that make up the backbone of regular, everyday speech. In fact, the connection between speech and beatboxing is so close that, in order to notate beatbox sounds, artists have used the International Phonetic Alphabet as a base for Standard Beatbox Notation.

Want to know how to make the classic kick drum sound? On the website Humanbeatbox.com, beatboxer Gavin Tyte explains how. First, he points out:

> In phonetics, the classic kick drum is described as a bilabial plosive (i.e., stop). This means it is made by completely closing both lips and then releasing them accompanied by a burst of air.

To punch up the sound, Tyte explains, you add a bit of lip oscillation, as if you were blowing a very short "raspberry." Step by step, in Tyte's words:

1. Make the "b" sound as if you are saying "b" from the word bogus.
2. This time, with your lips closed, let the pressure build up.
3. You need to control the release of your lips just enough to let them vibrate for a short amount of time.

The classic kick drum sound (represented as "b" in Standard Beatbox Notation) can be made as a voiced or voiceless version. Embellishments can be added: you can add on fricative sounds ("bsh," "bs," or "bf"), or combine the basic sound with a nasal sound ("bng," "bm," or "bn").

What sounds *really* impressive, though, is when a beatbox artist combines actual words with beatbox rhythms—it sounds as if the artist is simultaneously making speech sounds *and* non-speech sounds. But this is really a trick of the ear. It's not that the artist is making two sounds at the same time, but that he's creating a very convincing auditory illusion in which a *single* beatbox sound swings both ways, being heard both as a musical beatbox sound and as a speech sound. The illusion relies on what's known as the **phonemic restoration effect**. Scientists have created this effect in the lab by splicing a speech sound like "s" out of a word such as *legislature*, completely replacing the "s" sounds with the sound of a cough. Listeners hear the cough, but they also hear the "s" as if it had never been removed. This happens because, based on all the remaining speech sounds that really are there, the mind easily recognizes the word *legislature* and fills in the missing blanks (more on this in Chapter 7). In order for the illusion to work, though, the non-speech sound has to be acoustically similar to the speech sound. So, part of a beatboxer's skill lies in knowing which beatbox sounds can double as which speech sounds. Though many beatboxers have never taken a course in linguistics or psycholinguistics, they have an impressive body of phonetic knowledge at their command.

From a performance standpoint, skilled beatboxers display dazzling articulatory gymnastics. They keep their tongues leaping around their mouths in rapid-fire rhythms, and coordinate several parts of their vocal tracts all at the same time. But newbies to the art shouldn't be discouraged. It's certainly true that learning to beatbox takes many hours of practice. But when you think about it, the articulatory accomplishment is not all that different from what you learned to do as an infant mastering the sounds of your native language, and learning to put them all together into words. As you saw in Box 2.4, most infants spend quite a bit of time perfecting their articulatory technique, typically passing through a babbling stage beginning at about 5 months of age, in which they spend many hours learning to make human speech sounds. In the end, learning to beatbox may take no more practice than the many hours you were willing to put in learning how to talk—just think back to the hours you spent in your crib, taking your articulatory system out for a spin, and babbling endlessly at the ceiling.

phonemic restoration effect Auditory illusion in which people "hear" a sound that is missing from a word and has been replaced by a non-speech sound. People report hearing both the non-speech sound and the "restored" speech sound at the same time.

milliseconds) as it does for "b" in the word *ban*, we say the oral stop is a **voiced** one. When the vibration happens only at somewhat of a lag (say, more than 20 milliseconds), we say that the sound is **unvoiced** or **voiceless**. This labeling is just a way of assigning discrete categories to what amounts to a continuous dimension of **voice onset time** (**VOT**), because in principle, there can be any degree of voicing lag time after the release of the articulators.

WEB ACTIVITY 4.5

The phonetics of beatboxing Here you'll see some skilled beatboxers in action and learn more about the phonetics of beatboxing by watching tutorials on producing sounds such as the classic kick drum and the brushed snare.

You might have noticed that all of the consonants listed in Figure 4.6 end up being different phonemes. That is, it's possible to take any two of them and use them to create minimal pairs, showing that the differences between these sounds lead to differences in meaning, as we saw in Table 4.1. But that table shows you only a **phonemic inventory** of English sounds, not the full range of how these sounds are produced, in all their glorious allophonic variety, when each phoneme trots out its full wardrobe.

Phonemes versus allophones: How languages carve up phonetic space

Now that you have a sense of the dimensions along which sounds vary, I owe you a convincing account of why the differences between phonemes are often no bigger than the differences between allophones. (I'm talking here about their differences in terms of pure sound characteristics, not your mental *representations* of the sounds.)

Let's first talk about the differences in the "t" sounds in *tan* and *Stan*. Remember that extra little burst of air when you said *tan*? That actually comes from a difference in voice onset time. That is, there's an extra-long lag between when you release your tongue from your alveolar ridge and when your vocal folds begin to vibrate, perhaps as long as 80 milliseconds (ms). You get that extra puff because more air pressure has built up inside your mouth in the meantime. Unvoiced oral stops with a longer voice onset time are called **aspirated stops**, and these are the sounds that "pop" if you get too close to a microphone without a pop filter. Following standard notation, we'll use slightly different symbols for aspirated stops—for example, p^h, t^h, and k^h (with superscripts) to differentiate them from **unaspirated stops** t, d, and k. From here on, I'll also follow the standard practice in linguistics, and instead of using quotation marks around individual sounds, I'll indicate whether I'm referring to phonemes by enclosing them in forward slashes (for example, /b/, /d/), while allophones will appear inside square brackets (for example, [t], [t^h]).

Now, notice that a difference in voice onset time is exactly the way I earlier described the distinction between the phonemes /t/ and /d/—sounds that are distinguished in minimal pairs and that cause sudden shifts of meaning (see Figure 4.7). The difference between /t/ and /d/ seems obvious to our ears. And yet we find it hard to notice a similar (and possibly even larger) difference in VOT between the [t] and [t^h] sounds in *Stan* and *tan*. If we become aware of the difference at all, it seems extremely subtle. Why is this? One possible explanation might be that differences at some points in the VOT continuum are inherently easier to hear than distinctions at other points (for example, we might find that the human auditory system had a heightened sensitivity to VOT differences between 10 and 30 ms, but relatively dull perception between 30 and 60 ms). But another possibility is that our perceptual system has become tuned to sound distinctions differently, depending on whether those distinctions are allophonic or phonemic in nature. In other words, maybe what we

voiced Describes a sound that involves vibration of the vocal folds; in an oral stop, the vibration happens just about simultaneously with the release of the articulators (say, within about 20 milliseconds) as it does for "b" in the word *ban*.

unvoiced (voiceless) Describes a sound that does not involve simultaneous vibration of the vocal folds; in a voiceless stop followed by a vowel, vibration happens only after a lag (say, more than 20 milliseconds).

voice onset time (VOT) The length of time between the point when a stop consonant is released and the point when voicing begins.

phonemic inventory A list of the different phonemes in a language.

aspirated stop An unvoiced oral stop with a long voice onset time and a characteristic puff of air (aspiration) upon its release; an aspirated stop "pops" when you get too close to a microphone without a pop filter. Aspirated stop sounds are indicated with a superscript: p^h, t^h, and k^h.

unaspirated stop An unvoiced oral stop without aspiration, produced with a relatively short VOT.

"hear" isn't determined only by the objective acoustic differences between sounds, but also by the *role* that sounds play within the language system.

As it happens, different languages don't necessarily put phoneme boundaries in the same places; they can differ quite dramatically in how they carve up the sound space into phonemic categories. These distinctions make it possible for us to study whether our perception of sounds is influenced by how a language organizes sound into these categories. Speakers of Thai, for example, distinguish not only between the voiced and unvoiced phonemes /t/ and /d/, but also between the aspirated voiceless phoneme /t^h/ (as in *tan*) and its unaspirated version /t/ (as in *Stan*). What this means is that if you're speaking Thai, whether or not you aspirate your stops makes a difference to the meaning. Slip up and aspirate a stop by mistake—for example, using /t^h/ rather than /t/—and you've uttered a word that's different from the one you'd intended.

On the other hand, Mandarin, like English, has only two phonemic categories. But unlike English, Mandarin speakers make a meaningful distinction between voiceless aspirated and unaspirated sounds rather than voiced and voiceless ones. To their ears, the differences between /t/ and /t^h/ is painfully obvious, corresponding to different phonemes, but they struggle to "hear" the difference between [t] and [d].

Looking across languages, it's hard to make the case that either the difference between voiced and voiceless sounds or the difference between aspirated and unaspirated sounds is *inherently* more obvious. Different languages latch on to different distinctions as the basis of their phonemic categories. This becomes all the more apparent when you consider the fact that languages differ even in terms of which dimensions of sound distinction they recruit for phonemic purposes, as we saw in Chapter 3.

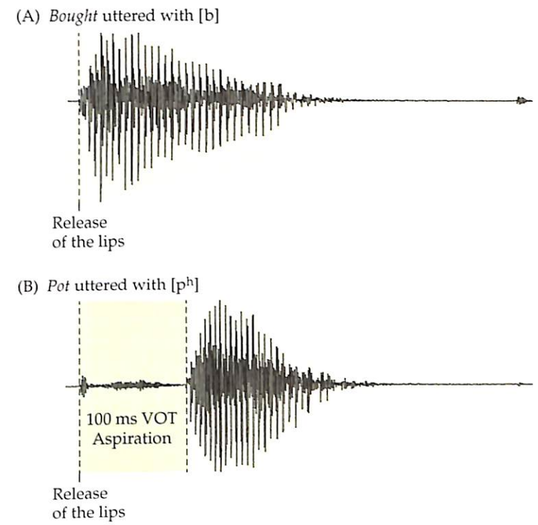


Figure 4.7 Waveforms for the words *bought* (A) and *pot* (B). *Bought* is uttered with a [b] sound at the beginning of the word (at a voice onset time of 0 ms), so that phonation (vocal fold vibration) occurs simultaneously with the release of the lips. *Pot* is uttered with a [p^h] sound, with a lag of 100 ms occurring between the release of the lips and the beginning of phonation. (Courtesy of Suzanne Curtin.)

For example, in English, whether or not a vowel is stretched out in time is an allophonic matter (Box 4.3). Vowels tend to be longer, for instance, just before voiced sounds than voiceless ones, and can also get stretched out for purely expressive purposes, as in "no waaay!"—note that *waaay* is still the same word as *way*. There's no systematic phonemic distinction between long and short vowels. But in some languages, if you replace a short vowel with a longer one, you'll have uttered a completely different word (for example, if you lengthen the vowel in the Czech word for *Sir*, you'll be addressing someone as *cheese*). In a similar vein, in Mandarin and various other languages described as "tone languages," the *pitch* on a vowel actually signals a phonemic difference. You might have just one sequence of vowels and consonants that will mean up to six or seven different words depending on whether the word is uttered at a high, low, or medium pitch, or whether it swoops upwards in pitch, whether the pitch starts high and falls, or whether the pitch rises and *then* falls. Needless to say, distinctions like these can be exasperatingly difficult to learn for speakers of languages that don't use tone phonemically. And, as you also saw in Chapter 3, there's evidence that different brain processes are involved in using these dimensions of sound, depending on the role they play in the language, with tonal differences on words eliciting more left-

WEB ACTIVITY 4.6

Phonemic distinctions across languages In this activity, you'll get a sense of how difficult it is to "hear" what are phonemic distinctions in other languages but allophonic in English.

BOX 4.3
Vowels

Unlike consonants, vowels are all made with a relatively unobstructed vocal cavity that allows the air to pass fairly freely through the mouth. Their various sounds are accomplished by shaping the mouth in different ways and varying the placement of the tongue. Interestingly, our perceptual systems tend not to be as categorical when hearing vowels as they are when perceiving consonants, and we're usually sensitive to even small graded differences among vowels that we'd lump into the same category—though there is evidence that experience with a particular language does have an effect on perception.

Vowels are normally distinguished along the features of **vowel height** (which you can observe by putting a sucker on your tongue while saying different vowels), **vowel backness**, **lip rounding**, and **tenseness**.

English has an unusual number of vowel sounds; it's not uncommon for languages to get by with a mere five or so. Only a couple of vowels ever occur in English as **diphthongs**, in which the vowel slides into an adjacent glide (as in the words *bait boat*). Below are the IPA symbols for the English vowel sounds, with examples of how they appear in words in a standard American dialect (note their uneasy relationship to English orthography):

i	beet	u	boot
ɪ	bit	ʊ	book
eɪ	bait	ow	boat
ɛ	bet	ɔ	bought
æ	bat	ɑ	cot
ə	the	ʌ	but
aj	bite	ɔj	boy
aw	bout		

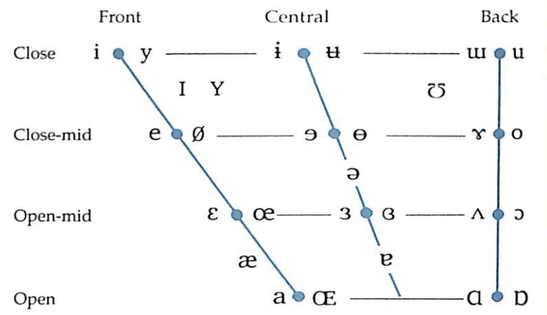


Figure 4.8 A vowel chart, a graphic illustration of the features of vowels, including English vowels and vowels found in other languages. When symbols are in pairs, the one to the right is the rounded version. Diphthongs like eɪ are not marked in this chart but represent transitions between vowels.

The features of the English vowels, along with others that don't occur in English, can be captured graphically in a vowel chart such as the one in Figure 4.8.

vowel height The height of your tongue as you say a vowel; for example, *e* has more vowel height than *a*.

vowel backness The amount your tongue is retracted toward the back of your mouth when you say a vowel.

lip rounding The amount you shape your lips into a circle; for example, your lips are very rounded when you make the sound for *w*.

tenseness A feature of vowels distinguishing "tense" vowels such as those in *beet* and *boot* from "lax" vowels such as those in *bit* and *put*.

diphthong A sound made when the sound for one vowel slides into an adjacent glide in the same syllable, as in the word *ouch*.

hemisphere activity among Mandarin speakers, but more right-hemisphere activity among English speakers.

I've just shown you some examples where other languages have elevated sound distinctions to phonemic status, whereas the same distinctions in English have been relegated to the role of mere sound accessories. The reverse can be true as well. For instance, the English distinction between the liquid sounds /r/ and /l/ is a phonemic one; hence, it matters whether you say *rice for Lent* or *lice for rent*. But you'll probably have noticed that this distinction is a dastardly one for new English language learners who are native speakers of Korean or

Japanese—they are very prone to mixing up these sounds. This is because the difference between the two sounds is an *allophonic* one in Korean and Japanese, and speakers of these languages perceive the difference between the two sounds as much more subtle than do native English speakers.

All of this goes to show that when it comes to how we perceive speech, we aren't just responding to the actual physical sounds out in the world. The way in which we hear sounds also has a lot to do with the structure our minds impose on sounds of speech. These mental structures can have dramatic effects in perceptually boosting some sound distinctions and minimizing others. We no longer interpret distinctions among sounds as gradual and continuous. This is actually a good thing, because it allows us to ignore many sound differences that aren't meaningful. For example, your typical English voiced [ba] sound might occur at a VOT of 0 ms, and your typical unvoiced [pʰa] sound might be at 60 ms. But your articulatory system is simply not precise enough to always pronounce sounds at the same VOT (even when you are completely sober); in any given conversation, you may well utter a voiced sound at 15 ms VOT, or an unvoiced sound at 40 ms. But your mind is very good at ignoring this articulatory slippage. What you know about the sound structures of your language imposes sharp boundaries, so you categorize sounds that fall within a single phoneme category—even if they're different in various ways—as the same, whereas sounds that straddle phoneme category boundaries clearly sound different. This way of perceiving sounds is called **categorical perception**, and it's quite a handy perceptual strategy.

categorical perception A pattern of perception where changes in a stimulus are perceived not as gradual, but as falling into discrete categories. Here, small differences between sounds that fall within a single phoneme category are not perceived as readily as small differences between sounds that belong to different phoneme categories.

forced choice identification task An experimental task in which subjects are required to categorize stimuli as falling into one of two categories, regardless of the degree of uncertainty they may experience about the identity of a particular stimulus.

WEB ACTIVITY 4.7

Categorical versus continuous perception In this activity, you'll listen to sound files that will allow you to compare perception of voiced and unvoiced consonants to the perception of pitch and volume.

To get a sense of the usefulness of categorical perception in real life, it's worth thinking about some of the many examples in which we don't carve the world up into clear-cut categories. Consider, for example, the objects in Figure 4.9. Which of these objects are cups, and which are bowls? It's not easy to tell, and you may find yourself disagreeing with some of your classmates about where to draw the line between the two (in fact, that line might readily shift depending on whether these objects are filled with coffee or soup). What's interesting is that this sort of disagreement is not likely to arise when it comes to consonants that hug the dividing line between two phonemic categories.

Such lack of disagreement is a hallmark of categorical perception, and it's been amply demonstrated in many experiments. One common way to test for categorical perception is called a **forced choice identification task**. The strategy is to have people listen to many examples of speech sounds and indicate which one of two categories each sound represents (for example, /pa/ versus /ba/). The speech sounds are created in a way that varies the VOT in small increments—for example, participants might hear examples of each of the two sounds at 10-ms increments, all the way from –20 ms to 60 ms. (A *negative* VOT value means that vocal fold vibration begins even before the release of the articulators.)

If people were paying attention to each incremental adjustment in VOT, you'd find that at the extreme ends (i.e., at –20 ms and at 60 ms), there would be tremendous agreement about whether a sound represents a /ba/ or a /pa/, as seen

Figure 4.9 Is it a cup or a bowl? The category boundary isn't clear, as evident in these images, inspired by a classic experiment by linguist Bill Labov (1972). In contrast, the boundary between different phonemic categories is quite clear for many consonants.

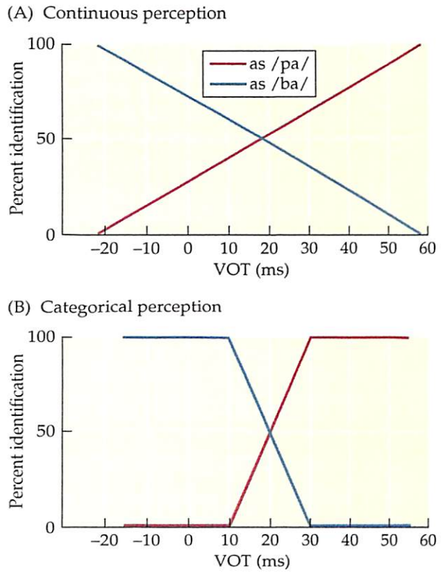


Figure 4.10 Idealized graphs representing two distinct hypothetical results from a phoneme forced-choice identification task. (A) Hypothetical data for a perfectly continuous type of perception, in which judgments about the identity of a syllable gradually slide from /ba/ to /pa/ as VOT values increase incrementally. (B) Hypothetical data for a sharply categorical type of perception, in which judgments about the syllable's identity remain absolute until the phoneme boundary, where the abruptly shift. Although there's some variability depending on the specific tasks and specific sounds, most consonants that represent distinct phonemes yield results that look more like (B) than (A).

in **Figure 4.10A**. In this hypothetical figure, just about everyone agrees that the sound with the VOT at –20 ms is a /ba/, and the sound with the VOT at 60 ms is a /pa/. But, as also shown in Figure 4.10A, for each step away from –20 ms and closer to 60 ms, you see a few more people calling the sound a /pa/.

But when researchers have looked at people's responses to forced choice identification tasks, they've found a very different picture, more like the graph in **Figure 4.10B**. People agree pretty much unanimously that the sound is a /ba/ until they get to the 20 ms VOT boundary, at which point the judgments flip abruptly. The upshot of all this is that when you're processing speech sounds, there's usually no inner mental argument going on about whether to call a sound /ba/ or /pa/. (The precise VOT boundary that separates voiced from unvoiced sounds can vary slightly, depending on the place of articulation of the sounds.)

What sound distinctions do newborns start with?

Put yourself in the shoes of the newborn, who is encountering speech sounds in all their rich variability for the first time (more or less: some aspects of speech sounds—especially their rhythmic properties—*do* make it through the uterus wall to the ears of a fetus, but many subtle distinctions among sounds will be encountered for the first time after birth). We've seen that adults don't pay equal attention to all sound distinctions—they pay special attention to those that signal differences between phonemic categories. But we've also seen that phoneme categories can vary from language to language, and that sound distinctions that are obvious to one language group may be more elusive to another. Clearly, these distinctions have to be learned to some extent. So what is a newborn baby noticing in sounds? Given that she's unlikely to have formed categories such as /p/ and /b/, since these categories are somewhat language-specific, does this mean that she's paying attention to *every* possible way in which sounds might vary in their pronunciation? Remember that sounds can vary along a number of different dimensions, with incremental variation possible along any of these dimensions. Let's suppose that babies are perceiving continuously rather than categorically (see Figure 4.9) for any of these sound dimensions. In that case, the sound landscape for babies would be enormously cluttered—where adults cope with several dozen categories of speech sounds, babies might be paying attention to hundreds of potential categories.

It takes some ingenuity to test for categorical perception in newborns. Once again, you can't give these miniature humans a set of verbal instructions and get back a verbal response that will tell you whether they are perceiving the difference between certain sounds. You're stuck making do with behaviors that are within the reach of your average newborn—which, admittedly, are not a lot. Faced with a newborn whose behavioral repertoire seems limited to sleeping, crying, sucking, and recycling body wastes, a researcher might be forgiven for feeling discouraged. It turns out, though, that one of these behaviors—sucking—can, in the right hands, provide some insight into the infant's perceptual processes. Babies suck to feed, but they also suck for comfort, and if they happen to have something in their mouths at the time, they suck when they get excited. And, as may be true for all of us, they tend to get excited at a bit of novelty.

By piecing these observations together, Peter Eimas and his colleagues (1971), pioneers in the study of infant speech perception, were able to design an experimental paradigm that allows researchers to figure out which sounds babies are perceiving as the same, and which they're perceiving as different. The basic premise goes like this: If babies are sucking on a pacifier while hearing speech sounds, they'll tend to suck vigorously every time they hear a new sound. But if they hear the same sound for a long period of time, they become bored and suck with less enthusiasm. This means that a researcher can cleverly rig up a pacifier to a device that measures rate of sucking, and play Sound A (say, [pa]) over and over until the baby shows signs of boredom (that is, the baby's sucking slows down). Once this happens, the researcher can then play Sound B (say, [p^{h}a]). If the baby's sucking rate picks up, this suggests the baby has perceived Sound B as a different sound. If it doesn't, it provides a clue that the baby, blasé about the new sound, hasn't perceived it as being any different from the first one (see **Method 4.2** for details of this approach). When it comes to testing for categorical perception then, if babies perceive speech sounds categorically, they should be oblivious to differences between certain sounds but acutely sensitive to differences between other sounds that fall on different sides of a critical boundary. On the other hand, if they're perceiving continuously, then they should *always* hear Sound B as different, and should increase their sucking just about any time Sound B is introduced.

If we look at how babies perceive VOT, the experiments show clear evidence of categorical perception in newborns, so it appears that the youngest humans don't treat all sound distinctions in the same way. Their rate of sucking goes up when two sounds straddle a VOT boundary of about 25 ms, but otherwise they seem oblivious to differences in VOT. This boundary is very similar to the adult dividing line for English voiced and voiceless sounds. What this means is that the sound landscape comes pre-carved to some extent; upon birth, babies aren't faced with the massive task of considering *every* possible difference in sound as being potentially meaningful when it comes to signaling differences between phonemic categories. Some sound distinctions are more privileged than others right off the bat.

When researchers first discovered that babies emerge from the womb with certain pre-set boundaries that happen to line up with the VOT boundaries distinguishing English voiced and voiceless sounds, this generated some excited speculation. Some researchers suggested that children come innately equipped with a set of inborn phonetic categories that are commonly used by languages. But this line of thinking quickly ran into a wall. First of all, Patricia Kuhl and James Miller (1975) devised a clever experiment to study the perception of consonants by chinchillas—which, while adorable, are not known for their linguistic skills, and certainly don't ever produce speech, so it's doubtful that they would be born innately prepared for it. Kuhl and Miller found that

METHOD 4.2

High-amplitude sucking

The high-amplitude sucking method allows researchers to peer into the minds of babies who, due to their tender age, understandably have a limited repertoire of behaviors. It's based on the premise that infants will naturally suck on objects in their mouths when they are excited by hearing a new sound. Throughout the experiment, the baby participant sucks on a pacifier that contains electrical circuitry that measures the pressure of each sucking motion so that the rate of sucking can be constantly tracked. The pacifier is held in place by an assistant, who wears headphones and listens to music to block out the experimental sounds, so there's no possibility of sending any inadvertent signals to the baby.

Infants tend to suck with gusto when they hear a new, interesting sound anyway, but in order to get the strongest connection possible between a new stimulus sound and this sucking behavior, researchers have an initial conditioning phase built into the session. During this phase, a new sound is played every time the baby sucks on the pacifier at a certain rate: no vigorous sucking, no terrific new sound. Babies quickly learn to suck to hear the stimulus.

Once the baby has been trained to suck to hear new sounds, researchers play the first of a pair of stimulus sounds—let's say a [pa] sound—over and over. Once the baby's interest lags, the sucking rate goes down. This is an example of **habituation**. When the baby's sucking rate dips below a criterion level that's been previously established, the second sound of the pair is played, and the dependent measure is the sucking rate of the baby *after* the presentation of this new sound. If the baby begins to suck eagerly again, it's a good sign that she perceives the second sound as different from the first.

This method can be adapted to measure which of two kinds of stimuli infants prefer. For example, if you wanted to know whether infants would rather listen to their own language or an exotic tongue, you could set up your study in a way that trains babies to suck slowly or not at all in order to hear one language, and to suck hard and fast in order to hear the other (being sure to counterbalance your experimental design to make sure that an equal number of babies are trained to suck slowly versus quickly for the native language).

As you might imagine, when working with extremely new babies, there can be quite a lot of data loss. Often babies are too tired, too hungry, or just too ornery to pay much attention to the experimental stimuli, so a large number of participants must be recruited for this method. The method works quite well for infants up to about 4 months of age, after which point the pastime of sucking begins to lose some of its appeal for infants, and they're less likely to keep at it for any length of time. Luckily, at around this age the head-turn preference paradigm becomes an option for testing babies' perception of speech.

habituation Decrease in responsiveness to a stimulus upon repeated exposure to the stimulus.

these small, furry mammals also perceived consonants categorically, along a VOT boundary very similar to the one found for humans (see **Box 4.4**). This result has since been replicated in some of our closer relatives, such as macaque monkeys, and in much more distant animal relatives such as birds.

There's a second problem with the notion that categorical perception in human newborns reflects innate preparation for linguistic sounds: it turns out that many *non*-speech sounds are perceived categorically as well—not just by humans but also by animals that are *very* distant from us on the evolutionary family tree, such as crickets and frogs. So it looks as if the process of amplifying some sound distinctions while minimizing others is a very general property of the auditory system across species. Though it has a certain usefulness for perceiving speech, it doesn't seem to be intrinsically related to speech.

An especially telling demonstration of the parallels in perception of speech and non-speech sounds comes from experiments that use non-speech sounds to mimic some of the properties of human speech sounds. Remember that VOT

BOX 4.4

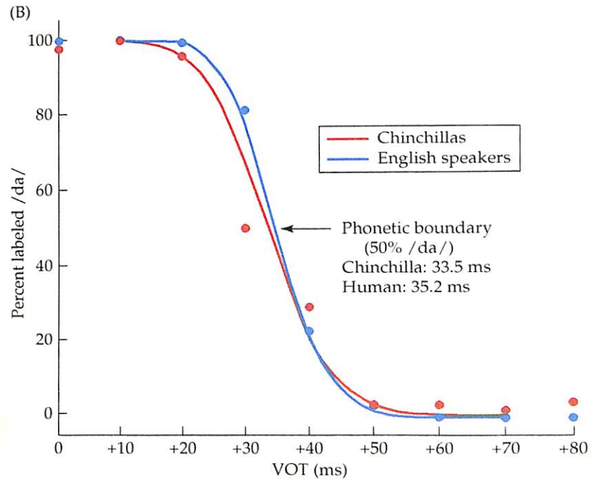
Categorical perception in chinchillas

How can you study speech perception in animals? As they do with babies, scientists have to find a way to leverage behaviors that come naturally to animals, and incorporate these behaviors into their experiments. Speech scientists Patricia Kuhl and James Miller (1975) took advantage of the fact that in many lab experiments, animals have shown that they can readily link different stimuli with different events, and that they can also learn to produce different responses to different stimuli in order to earn a reward.

In Kuhl and Miller's study, chinchillas (**Figure 4.11A**) heard various speech sounds as they licked a drinking tube to get dribbles of water. When the syllable /da/ was played (with a VOT of 0 ms), it was soon followed by a mild electric shock. As would you or I, the chinchillas quickly learned to run to the other side of the cage when they heard this sound. On the other hand, when the syllable /ta/ came on (with a VOT of 80 ms), there was no electric shock, and if the chinchillas stayed put and continued to drink, they were rewarded by having the water valve open to allow a stronger flow of water. In this way, the researchers encouraged the chinchillas to link two very different events to the different phonemic categories, a distinction the chinchillas were able to make.

The next step was to systematically tweak the voice onset times of the speech stimuli in order to see how well the chinchillas were able to detect differences between the two sounds at different points along the VOT continuum. Even though the animals had only heard examples of sounds at the far ends of the VOT spectrum, they showed the same tendency as human babies and adults do—that is, to sort sounds into clear-cut boundaries (see Figure 4.10)—and the sharp boundary between categories occurred at almost the same VOT as was found for humans (33.5 ms versus 35.2 ms; see **Figure 4.11B**).

Figure 4.11 (A) A chinchilla; these animals are rodents about the size of a squirrel. They are a good choice for auditory studies because the chinchilla's range of hearing (20–30 kHz) is close to that of humans. (B) Results from Kuhl and Miller's categorical perception experiment, comparing results from the animals and human adults. The graph shows the mean percentage of trials in which the stimulus was treated as an instance of the syllable /da/. For humans, this involved asking the subjects whether they'd heard a /da/ or /ta/ sound; for chinchillas, it involved seeing whether the animals fled to the other side of the cage or stayed to drink water. (After Kuhl and Miller, 1975.)



is a measure of the time between the release of the articulators and the beginning of voicing (that is, vibration of the vocal folds). A slightly more abstract way of looking at it is that the perception of VOT is about perceiving the relative timing of two distinct events. This scenario can easily be recreated with non-

ABX discrimination task A test procedure in which subjects hear two different stimuli followed by a third which is identical to one of the first two. The subjects must then decide whether the third stimulus is the same as the first or the second.

speech stimuli, simply by putting together two distinct sounds and playing around with their relative timing.

Researcher David Pisoni (1977) created a set of stimuli by using two distinct tones and varying the number of milliseconds that elapsed between the onsets of the two tones, much as was done in previous VOT experiments—we can call this "tone onset time," or TOT. He then tested to see whether there was a certain window across which people would be especially sensitive to TOT differences. For instance, people might hear two stimuli, Stimulus A being two tones whose onsets were separated by 20 ms, and Stimulus B being two tones separated by 30 ms. The people would then have to judge whether a third stimulus (in which the two tones were separated by, say, 30 ms) was the same as Stimulus A or Stimulus B. The idea behind this task, known as an **ABX discrimination task**, is that if people can readily perceive the difference between the two sound pairs, they'll be reliable at identifying whether the third sound pair is identical to the first or second. On the other hand, if they don't perceive the difference between them, then they'll be randomly guessing as to the identity of the third sound pair.

What Pisoni found was that people were especially good at distinguishing between stimuli right around a TOT of 20 ms. For example, the above pair of stimuli, sound pairs with TOTs of 20 ms and 30 ms, would be perceived as distinct by many of the subjects. But if people heard a pair of stimuli with sounds separated by 40 ms and 50 ms, they were much less likely to perceive them as different. The same was true for a pair of simultaneously produced sounds (0 ms TOT) and a sound pair 10 ms apart. In other words, the TOT boundary for optimal perception of differences was strikingly similar to the boundary for *voice* onset time of speech sounds. Pisoni suggested that differences at about the 20 ms boundary for both speech and non-speech sounds are easy to notice because this is the point at which the auditory system is able to detect that two events occurred at different times. If the time between two events is any shorter, it becomes hard to perceive that they didn't occur at the same time. The limits of the auditory system make the 20 ms mark a point at which stimuli naturally divide up into categories of simultaneous versus non-simultaneous pairs of sound events.

Clearly, the overall evidence from categorical perception scores no points for the hypothesis that babies come preinstalled with probable speech categories. Instead, it supports the notion that a language like English is being opportunistic about where it carves phonemic categories—it appears to be shaping itself to take advantage of natural perceptual biases of the auditory system.

Still, not all languages take advantage of the natural places to carve up phonemic categories that the auditory system so conveniently offers up. As we saw, languages like Mandarin opt not to distinguish between phonemes at the "natural" boundaries, placing phonemic boundaries elsewhere instead. Since babies are obviously able to grow into Mandarin-speaking adults, there must be enough flexibility in their perceptual systems to adapt to the categories as defined by their particular language. What changes in the perceptual life of an infant as she digests the sounds of the language around her?

Quite a bit of research has shown that babies start off noticing a large number of distinctions among sounds, regardless of whether the languages they'll eventually speak make use of them to mark phonemic distinctions. For example, all babies, regardless of their native languages, start off treating voiced and unvoiced sounds as different, and the same goes for aspirated versus unaspirated sounds. As they learn the sound inventory of their own language, part of their job is to learn which variations in sounds are of a deep, meaning-

changing kind, and which ones are like wardrobe options. Eventually, Mandarin-hearing babies will figure out that there's no need to separate voiced and voiceless sounds into different categories, and they will downgrade this sound difference in their auditory attention. (Here's a workable analogy to this attentional downgrading: presumably, you've learned which visual cues give you good information about the identity of a person, and which ones don't, so you pay more attention to those strongly identifying cues. So, you might remember that you ran into your co-worker at the post office, but have no idea what she was wearing at the time.) Unlike the Mandarin-hearing babies, who "ignore" voicing, English-hearing babies will learn to "ignore" the difference between aspirated and unaspirated sounds, while Thai-hearing babies will grow up maintaining a keen interest in both of these distinctions.

It can be a bit humbling to learn that days-old babies are good at perceiving sound differences that you strain to be aware of. There's a large body of research (pioneered by Janet Werker and Richard Tees, 1984) that now documents the sounds that newborns tune in to, regardless of the language their parents speak. Unlike many of you, these tiny bundles of joy can easily cope with exotic sound distinctions, including these: the subtle differences among Hindi stops (for instance, the difference between a "regular" English-style /t/ sound and one made by slightly curving the tip of your tongue back as you make the sound); Czech fricatives (for instance, the difference between the last consonant in *beige* and the unique fricative sound in *Dvorak*); and whether a vowel has a nasal coloring to it, a distinctive feature in French, important for distinguishing among vowels that shift meaning. At some point toward the end of their first year, babies show evidence of having reorganized their perception of sounds. Like adults, they begin to confer special status on those distinctions that sort sounds into separate phonemic categories of the language they're learning.

WEB ACTIVITY 4.8

Distinct sounds for babies

In this activity you'll listen to some non-English sound distinctions that newborn babies can easily discriminate.

4.4 Learning How Sounds Pattern

The distribution of allophones

In the previous section, we saw that babies start by treating many sound distinctions as potentially phonemic, but then tune their perception in some way that dampens the differences between sounds that are non-phonemic in the language they're busy learning. So, a Mandarin-hearing baby will start out being able to easily distinguish between voiced versus unvoiced sounds, but will eventually learn to ignore this difference, since it's not a phonemic one.

But this raises a question: How do infants *learn* which sounds are phonemic, and therefore, which differences are important, and which can be safely ignored? You and I know that voicing is a distinctive feature partly because we recognize that *bat* and *pat* are different words with different meanings. But remember that babies are beginning to sort out which sound differences are distinctive as early as 6 months of age, at a time when they know the meanings of very few words (a topic we'll take up in Chapter 5). If infants don't know what *bat* and *pat* mean, or even that *bat* and *pat* mean different things, how can they possibly figure out that voicing (but not aspiration) is a distinctive feature in English?

As it happens, quite aside from their different roles in signaling meaning differences, phonemes and allophones *pattern* quite differently in language. And, since babies seem to be very good at noticing statistical patterns in the