

Chapter 1

Knowing What to Expect

In This Chapter

- ▶ Clarifying what NaturallySpeaking can do
 - ▶ Figuring out what NaturallySpeaking can't do
 - ▶ Selecting the right NaturallySpeaking product
-

Voice recognition is used in places like cars, hospitals, and legal offices. Yet, some people are still skeptical about software that enables you to dictate to your computer and get a transcription of what you said. People think it's very cool, but they secretly wonder if it really works.

It works. (And it *is* really cool.) Right out of the box, today's Dragon NaturallySpeaking reports 99 percent accuracy. I bet that's a score you'd like for your own personal output. Well then, read this book and dive right in. You'll be rewarded with higher productivity and hands-free computing.

Sections of this book were written by dictating them into Dragon NaturallySpeaking. It was a lot of fun, and I predict that you will also find NaturallySpeaking to be useful and fun — if you approach it with the appropriate expectations.

Clarifying What NaturallySpeaking Can Do for You

Something about dictating to a computer awakens all kinds of unrealistic expectations in people. If you expect it to serve you breakfast in bed, you're out of luck. I didn't write this book by saying, "Computer, write a book about NaturallySpeaking." I had to dictate it word for word, just as I would have had to type it word for word if I didn't have NaturallySpeaking.

So what are realistic expectations? Think of NaturallySpeaking the way that you think about your keyboard and mouse. It's an input device for your computer, not a brain transplant. It doesn't add any new capabilities to your

computer beyond deciphering your spoken words into text or ordinary PC commands. If you say, “**Go make me a sandwich,**” NaturallySpeaking will dutifully type “go make me a sandwich” into whatever word-processing application happens to be open.

Just because your computer can understand what you say, don’t expect it to understand what you mean. It’s still just a computer, you know.

Here are five specific things you *can* expect to do with NaturallySpeaking, and where to look for more details about how to do them:

- ✔ **Browse the web.** If you use Internet Explorer (IE) (or Chrome or Firefox) and NaturallySpeaking together, you can cruise around the web without ever touching your keyboard or mouse. Pick a website from your Favorites menu, follow a link from one web page to another, or dictate a URL (web address) into the Address box, leaving your hands in your lap the whole time. See Chapter 13 for details.
- ✔ **Control your applications.** If you see the name on a menu, you can say click and the name and watch it happen — not just in NaturallySpeaking but in your other applications as well. If your e-mail program has a Check Mail command on its menu, then you can check your e-mail by saying a few words. Anything that your spreadsheet has on a menu becomes a voice command you can use. Ditto for hotkeys: If pressing some combination of keys causes an application to do something you want, just tell NaturallySpeaking to press those keys. See Part II.
- ✔ **Control your desktop.** Applications will start running just because you tell them to. Use your voice to open and close windows, switch from one open window to another, and drag and drop stuff from here to there. See Chapter 15.
- ✔ **Dictate into a digital recorder and let NaturallySpeaking transcribe it later.** You need NaturallySpeaking and a digital (or very good analog) recorder. See Chapter 11.
- ✔ **Write documents.** NaturallySpeaking is darn good at helping you write documents. You talk, and it types. If you don’t like what you said (or what NaturallySpeaking typed), tell NaturallySpeaking to go back and change it. You can give vocal instructions to make elements bold, italic, large, small, or set in a particular font. Chapters 3, 4, 5, and 6 explain what you need to know to write documents in the NaturallySpeaking DragonPad itself. If you want to dictate into Microsoft Word or Corel WordPerfect, see Chapter 9. For all other applications, see Chapter 8.

Now, I happen to think that’s plenty to get excited about. I can make my own sandwiches, thank you.

Figuring Out What NaturallySpeaking Can't Do

Even with NaturallySpeaking, your computer's capability to understand English is more limited than what you can reasonably expect from a human. People use a very wide sense of context to figure out what other people are saying. We know, for instance, what the teen behind the counter at Burger King means when he asks, "Wonfryzat?" (That's fast-food-employed teenspeak for "Do you want fries with that?") We'd be completely confused if that same teenager walked up to us at a public library and asked, "Wonfryzat?"

NaturallySpeaking figures things out from context, too, but only from the *verbal* context (and a fairly small verbal context at that). It knows that "two apples" and "too far" make more sense than "too apples" and "two far." But two- to three-word context seems to be about the extent of the software's powers. (I can't say exactly how far it looks for context, because Nuance, the manufacturer, is understandably pretty hush-hush about the inner workings of NaturallySpeaking.) It doesn't understand the content of your document, so it can't know that words like "Labradoodles" and "Morkies" are going to show up just because you're talking about dogs.

However, Dragon does use context in its analysis. NaturallySpeaking has collected millions of hours of dictation from customers and uses this data to better understand the language. A word receives a frequency score that shows how often the word appears and how often the word appears before or after other words. This data is then used to determine the best choice; for example, when you mention "too" or "two."



TIP

Try to speak in slightly longer utterances. The more you say in one breath the more data Dragon has. But be careful, try to say too much at one time and you will likely trail off or start to mumble as you lose concentration.

Consequently, you can't expect NaturallySpeaking to understand every form of speech that humans understand. In order to work well, it needs advantages like these:



REMEMBER

- ✓ **Familiarity:** Each person who dictates to NaturallySpeaking can train it so that NaturallySpeaking builds an individualized user model. So NaturallySpeaking can't transcribe the voice mail that other people leave for you.

Even though you no longer have to train Dragon, it still creates a speaker-dependent profile. Once the audio setup is complete, NaturallySpeaking is already making it unique to you. That's why you must always use your own profile.

- ✓ **Identification:** Each time you start dictating, you need to identify yourself so that NaturallySpeaking can load the right user model.

- ✔ **One user at a time:** NaturallySpeaking loads only one user model at a time, so it can't transcribe a meeting during which several people talk, even if it has user models for all of them.
- ✔ **Constant volume:** You can't plunk a microphone down in the middle of the room and then pace around while you dictate.
 - Wear a good microphone (like the one that comes with NaturallySpeaking) and position it the same way every time or sit in front of your computer and try to stay at the same distance from your PC's internal array microphone.

For the first time, NaturallySpeaking supports using a PC's internal mic.
 - Don't mumble or let your voice trail off.

Dragon does a pretty good job with accents, though, as long as you're consistent.
- ✔ **Reasonable background noise:** Humans may be able to understand you when your favorite drummer is blasting away on your speakers or the blow dryer is on. They may be reading your lips at least part of the time, and they can guess that you're probably saying, "Turn that thing down!" NaturallySpeaking lacks in the lip-reading department, as well as in the capability to make obvious situational deductions.
- ✔ **Reasonable enunciation:** You don't have to start practicing "Moses supposes his toeses are roses," but you do need to realize that NaturallySpeaking can't transcribe sounds that you don't make. See Chapter 16 for an in-depth discussion of this issue.
- ✔ **Standard turn-of-the-millennium English prose:** If you want to be the next James Joyce, stick to typing. You can have some fun by trying to transcribe Shakespeare or things written in other languages (I had fun in Chapter 25), but it isn't going to work very well (unless you do some really extensive training). On the other hand, NaturallySpeaking is just the thing for writing books, blog posts, reports, short stories, and letters to Mom.



Letting someone else use your profile will negatively affect your accuracy.

Selecting the Right NaturallySpeaking Product

NaturallySpeaking isn't a single product; it's a family of products. And like most families, some members are richer than others. Depending on the features you want, you can pay a hefty price for software. You get what you pay for.

In spite of their socioeconomic differences, this family gets along pretty well. All the products are based on the same underlying voice recognition system, so they create the same kinds of user files. This fact has two consequences for you as a user:

- ✔ All of the products are about equally accurate at transcribing your speech. However, the Legal version has specific legal terms included and the language models are more tuned to expect legal phrases. This holds true for the Medical version as well.
- ✔ Upgrading to a better version is easy.
You can start out with the inexpensive Home edition, test out whether you like this whole idea of dictation, and then move up to a full-featured version without having to go through training all over again.

Which edition is best for you depends on why you're interested in NaturallySpeaking in the first place. Are you a poor typist who wants to be able to create documents more quickly? A good typist who is starting to worry about carpal tunnel syndrome? A person who can't use a mouse or keyboard at all? A busy executive who wants to dictate into a recorder rather than sit in front of a monitor? Is price an important factor to you? Do you need NaturallySpeaking to recognize a large, specialized vocabulary? Do you want to create macros that enable you to dictate directly into your company's special forms?

The more features you want, the more you should expect to pay.

Expanding the use of speech recognition

Speech recognition software is entrenched in many private sector industries. Dragon NaturallySpeaking serves several industries, including the following:

- ✔ **Financial:** NaturallySpeaking helps financial people manage their paperwork and meet compliance requirements.
- ✔ **Legal and medical:** Transcription and documentation play a major role in keeping things moving in the legal and medical fields. NaturallySpeaking significantly cuts the time needed to produce various documents.
- ✔ **Insurance:** This one is self-explanatory. Anything that cuts down on paperwork in the insurance industry is clearly a public service.

The public sector uses Dragon NaturallySpeaking as follows:

- ✔ **Education:** It is well documented that NaturallySpeaking can help level the playing field for students who face learning challenges. Teachers can provide better learning experiences to all their students.

- ✔ **Accessibility:** NaturallySpeaking makes a major contribution to people who are challenged by the use of a keyboard or mouse. The software provides access to the web and opens up the world to people who might otherwise be denied digital access.
- ✔ **Public safety:** Dragon's capability to save time on paperwork frees up law enforcement professionals to do the work that keeps us safe.

The latest generation of the NaturallySpeaking family

The current generation of NaturallySpeaking, version 13, was released in the second half of 2014. In addition to the usual bug fixes and incremental improvements that you expect in a new version of an application, NaturallySpeaking 13 brings the following five major enhancements:

- ✔ **Improved accuracy:** Every facet of NaturallySpeaking works faster by cutting down on the time required for the program to recognize dictation and produce output. Nuance reports that NaturallySpeaking 13 is 15 percent more accurate than version 12 and doesn't require a training session.
- ✔ **Faster response time:** Response to commands is faster and makes dictating less about stopping and starting. You can pay attention to longer phrases with pauses in between.
- ✔ **Shortcuts for common commands:** Nuance has anticipated many of the common commands that you want to use and has created shortcuts for them.
- ✔ **No training to set up:** You don't have to spend time reading documents. Nuance has made this version a five-minute task.
- ✔ **A newly designed DragonBar:** Using the DragonBar is easier than ever before. It's streamlined and can be collapsed or opened at the click of a mouse. No extended menu needed either.

Here is the current lineup of some of the NaturallySpeaking products, with a few comments about their features:

- ✔ **NaturallySpeaking Home:** This entry-level edition is perfect for people who hate to type. It is as accurate as the more expensive editions, enables control of the Windows desktop, includes a Dictation Box for dictating into other applications, and enables you to browse the web by voice. The Home version includes Full Text Control for a number of applications and some Natural Language Commands for Word, WordPerfect, and OpenOffice Writer. It doesn't support Excel or playback of your own voice for corrections. The Home edition is perfect if

you're planning to dictate only the first draft of documents, which you then polish using a mouse and keyboard. This version probably isn't the best choice for people with physical disabilities.

- ✔ **NaturallySpeaking Premium:** Premium includes all the Home edition's features, plus a few extras. It enables you to select a piece of your document and play back your own dictation, a great feature when you're trying to correct a mistake that either you or NaturallySpeaking made 20 minutes ago. It also opens the possibility of dictating into a recorder (including a smartphone) and letting NaturallySpeaking transcribe it later. See Chapter 11.
- ✔ **NaturallySpeaking Professional:** This edition is the one to get if you're personally committed to using voice recognition for everything or if you're a manager planning to convert your entire office to NaturallySpeaking. The Professional edition has everything Nuance could think of to make your experience easier, including two great features: You can build your own specialized vocabularies and create your own voice command macros — or more precisely, the office geek can construct specialized vocabularies and commands tailored to match the way his office works, and then everyone else in the office can use them too.
- ✔ **NaturallySpeaking Legal and NaturallySpeaking Medical:** At heart, these two editions begin with the Professional edition, but Nuance has done some of the work that I describe about the office geek in the preceding bullet.
 - *Medical edition:* Comes out of the box knowing the names of obscure diseases, body parts, and pharmaceuticals. The Medical versions can be seen as their own family of products. There are versions for different types of healthcare providers and organizations. It is important to know that the Medical versions of Dragon are the only versions that will allow you to dictate into an electronic medical report (EMR or EHR) application.
 - *Legal edition:* Knows *amicus curiae*, *habeas corpus*, and a bunch of other Latin legal terminology that would make the Professional edition throw up its proverbial hands.
- ✔ **Dragon NaturallySpeaking for the Mac:** I cover the NaturallySpeaking Windows products in this book, but Nuance also has a collection of products for the Mac, including Dragon Dictate and Dragon Dictate Medical for Mac. If you know some Mac users, tell them to check these out. (For use with mobile devices, including the iPhone, iPod, or iPad, see Chapter 14.)

In addition to these off-the-shelf products, you can also have NaturallySpeaking installed on your office network. This corporate option goes beyond the scope of what I cover in this book. If you're interested in this option, contact Nuance directly. Training programs for your staff are also available.

Understanding Speech Recognition in NaturallySpeaking

Nuance has eliminated the training during set-up. But if you really want to make NaturallySpeaking blazing fast, you should train it to understand you and the special words and phrases you use. So why does NaturallySpeaking need to be customized before it understands your speech? The simple answer is that speech recognition is probably one of the hardest things your computer does. Humans may not think speech recognition is hard, but that's because they are good at it. LeBron James probably has trouble understanding why the rest of us think it's so hard to dunk a basketball.

This section explains why deciphering speech is hard for computers and how training NaturallySpeaking can overcome these difficulties. I hope that understanding these issues will give you confidence that the extra effort is worth it.

NaturallySpeaking comes out of the box not knowing anything about you. It has to work as well for a baritone with a Scottish accent as for a mezzo-soprano with a slight lisp. It needs time to figure out how you talk.



Training continues for as long as you keep using NaturallySpeaking. It makes mistakes, you correct them, and it learns. That's the process. It gets better and better the more frequently you use it.

The exact error rate depends on many factors: how fast your computer is, how much memory it has, how good your microphone is, how quiet the environment is, how well you speak, what sound card your computer has, and so on.



NaturallySpeaking is 99 percent accurate out of the box and it gets better as long as you keep correcting it. Don't be lazy.

What's so hard about recognizing speech, anyway?

If 3-year-olds can recognize and understand speech (other than the phrase “go to bed”), why is it so hard for computers? Aren't computers supposed to be smart?

Well, yes and no. Computers are very smart when it comes to brain-straining activities like playing chess and filling out tax returns, so you may think they'd be whizzes at “simple” activities like recognizing faces or understanding speech. But after about 50 years of trying to make computers do these simple things, programmers have come to the conclusion that a skill isn't simple just because humans master it easily. In fact, our brains and eyes and ears are chock-full of sophisticated sensing and processing equipment that still runs rings around anything we can design in silicon and metal.

We humans think it's simple to understand speech because all the really hard work is done before we become conscious of it. To us, it seems as if English words just pop into our heads as soon as people open their mouths. The unconscious (or preconscious) nature of the process makes it doubly hard for computer programmers to mimic. If we don't know exactly what we're doing or how we do it, how can we tell computers how to do it?

To get an idea of why computers have such trouble with speech, think about something that they're very good at recognizing and understanding: touch-tone phone numbers. Those blips and bloops on the phone lines are much more meaningful to computers than they are to people. Several important features make the phone tones an easy language for computers, as I discuss in the following list. English, on the other hand, is completely different:

- ✔ **The touch-tone “vocabulary” has only 12 “words” in it.** After you know the tones for the ten digits plus * and #, you're in. English, on the other hand, has hundreds of thousands of words.
- ✔ **None of the words sound the same.** On the touch-tone phone, the “1” tone is distinctly different from the “7” tone. But English has homonyms, such as *new* and *gnu*, and near homonyms, like *merrier* and *marry her*. Sometimes entire sentences sound alike: “The sons raise meat” and “The sun's rays meet,” for example.
- ✔ **All “speakers” of the language say the words the same way.** Push the 5 button on any phone, and you get exactly the same tone. But an elderly man and a 10-year-old girl use very different tones when they speak; and people from Great Britain, Canada, and the United States pronounce the same English words in very different ways.
- ✔ **Context is meaningless.** To the phone, a 1 is a 1 is a 1. How you interpret the tone doesn't depend on the preceding number or the next number. But in written English, context is everything. It makes sense to “go to New York.” But it makes no sense to “go two New York” or “go too New York.”

What's a computer to do?

In order to work effectively for you, a speech-recognition program like NaturallySpeaking needs to combine four vastly different areas of knowledge. It needs to know a lot about speaking in general, about the spoken English language in general, about the way your voice sounds, and about your word-choice habits.

How NaturallySpeaking knows about speech and English in general

Dragon NaturallySpeaking gets its general knowledge from the folks at Nuance, some of whom have spent most of their adult lives analyzing how English is spoken. NaturallySpeaking has been programmed to know in general what

human voices sound like, how to model the characteristics of a given voice, the basic sounds that make up the English language, and the range of ways that different voices make those sounds. It has also been given a basic English vocabulary and some overall statistics about which words are likely to follow which other words. (For example, the word *medical* is more likely to be followed by *miracle* than by *marigold*.)

How NaturallySpeaking learns about your voice

NaturallySpeaking learns about your voice by listening to you. During the training process, you read out loud some text selections that NaturallySpeaking has stored in its memory. Because it already knows the text that you're reading, NaturallySpeaking uses this time to model your voice and learn how you pronounce words.

NaturallySpeaking goes on learning about your voice every time you use it. When you correct a word or phrase that NaturallySpeaking has guessed wrong, NaturallySpeaking adjusts its settings to make the mistake less likely in the future.

Dragon learns even when you make keyboard edits or delete everything and start over, taking the net result of your dictation or typed text and using it to help improve its accuracy.

NaturallySpeaking learns in two ways:

- ✓ **The Language Model:** When you make edits with your keyboard, Dragon will update your Language Model and better understand which words you use and when. However, NaturallySpeaking will never add a word to your vocabulary unless you add it manually in the Vocabulary Editor or use the “**Spell That**” option to make your correction.
- ✓ **The Acoustic Model:** Dragon becomes more accurate by updating your Acoustic Model, which is how you sound and pronounce words. When you make corrections with your voice, Dragon updates both your Language Model and your Acoustic Model because it now has both the text and the audio associated with that text.



Although you can correct with your keyboard, Dragon becomes smarter when you correct with your voice. Either way, though, Dragon gets better!

How NaturallySpeaking learns your word-choice habits

Initially, NaturallySpeaking learns how you choose words from the Vocabulary Builder phase of training. It may seem as if NaturallySpeaking is just learning how you say some unusual words. But, in fact, Vocabulary Builder is worthwhile even if no new words are found, because NaturallySpeaking analyzes how frequently you use common words and which words are likely to be used in combination.

NaturallySpeaking comes out of the box knowing general facts about the frequency of English words, but Vocabulary Builder helps sharpen those models for your particular vocabulary.

For example, if you want to use NaturallySpeaking to write letters to your mother and you let it study your previous letters, NaturallySpeaking will learn that the names of your family members appear much more frequently than they do in general English text. It is then much less likely to misinterpret your brother Johan's name as *John* or *yawn*.

Onward to Customizing!

By enabling you to install NaturallySpeaking, your computer has taken on one of the hardest tasks a PC ever faces. It needs your help. If you train it with patience and persistence, and if you gently but firmly correct your NaturallySpeaking assistant whenever it makes a mistake, you'll be rewarded with a computer that takes your verbal orders and transcribes your dictation without complaint (and even without a coffee break, unless you need one).

But I read in a magazine . . .

Quite a few magazine and newspaper articles have been written about voice recognition in general, and Dragon NaturallySpeaking in particular. Almost all of them contain an example that's something like this: A guy says into a microphone, **"Send e-mail to Bob about Friday's meeting. Period. Bob, comma, glad you're going to be there. Period."**

As if by magic, an e-mail application opens, a message window appears, Bob's e-mail address is pulled out of an address book somewhere, "Friday's meeting" is entered on the subject line, and the following text is entered into the message body: **"Bob, glad you're going to be there."**

The example is completely legitimate, as you can see in Chapter 12. But you need to keep something in mind: If you say, **"Do the numbers on February's revenue receipts,"** your spreadsheet just sits there.

Like any good magician's trick, more appears to be happening than actually happens. The computer has not suddenly been granted intelligence that rivals that of a human. The Nuance programmers have created a handful of scripts for doing everyday tasks, like generating e-mail messages and entering new events into a calendar program. They've made the commands sound like the instructions you would give a personal assistant, and they've set things up so that a lot of similar-sounding commands produce the same result. But if you say, **"Zip a message off to Bob,"** nothing happens.

Impressive? Yes. Magic? Not quite. It's still just a computer.

