

## **How It Works**

---

People recognize speech effortlessly and automatically. How to reproduce this ability in a computer is a question that has occupied speech researchers for decades. Several advancements have made electronic speech recognition possible: better understanding of sound and language, the development of specialized mathematical techniques, and vast improvements in computer speed and memory.

### **Many Sources of Knowledge**

---

Alexander Graham Bell tried to make human speech visible back in 1875. His wife, Mabel, had been deaf since age four, and he sought to create a machine that would generate pictures of the different frequencies in speech sounds. Mabel and other deaf people, Bell thought, might be able to understand speech by

looking at the graphs drawn by his machine. While experimenting he accidentally connected one of the wires to the wrong part of his apparatus. Sound unexpectedly came out of the microphone—Bell had invented the telephone. Bell did later get his machine to generate sound pictures, but the graphs proved too complex for humans to read as speech.

Bell's discovery is one of many that have made speech recognition a reality. Today's speech recognition programs do in fact break down speech into frequencies, as Bell sought to accomplish. They employ many other techniques and information sources as well. Speech recognition programs rely on knowledge about what sounds people make, which differs depending on the language spoken. Japanese, for example, has about 120 possible syllables, while English has more than 10,000. Speech software also incorporates information on sentence structure to help distinguish between words like "to," "too," and "two." NaturallySpeaking and programs like it also learn as you use them. They adapt to the sound of your voice and learn what words and phrases you use most often.

Today's PC has processing power and memory that was just a futuristic dream to early speech researchers. As Dragon Systems co-founder James Baker writes in the foreword to this book, early mainframe computers would take up to an hour to perform the millions of calculations required to recognize a single sentence. Processing power and memory capacity have since shot upwards, while their cost has plummeted. These engineering innovations are essential in making speech recognition practical on your desktop.

## How NaturallySpeaking Works

---

NaturallySpeaking recognizes your speech by using both the sound of your voice and a statistical model of what words tend to go together. Both pieces of information are vital for NaturallySpeaking to achieve acceptable levels of accuracy.

### The Sound of Your Voice

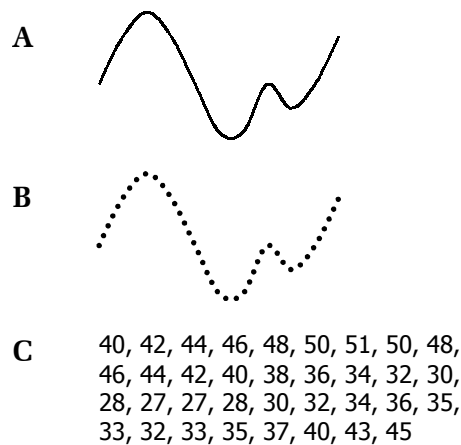
As you speak, your vocal folds vibrate and resonate in your chest and throat, creating the unique sound of your voice. This vibration travels through the air like waves moving outward from a stone dropped in a pond. The air vibration reaches your

listener's ears, her eardrum vibrates, and her brain interprets this vibration as speech, figuring out your words instantly and unconsciously.

The computer "hears" your voice through a microphone. Microphones have an electrical element sensitive to vibration—an artificial eardrum, in a sense. As the microphone element vibrates, it creates an electrical signal that changes just as fast as your vocal cords vibrated.

The sound card in your computer measures this changing electrical signal, assigning numbers to the signal more than 20,000 times per second. These measurements are so frequent that they give a quite accurate picture of the shape of the electrical changes. This process is called analog-to-digital conversion (Figure 19-1).

**Figure 19-1**



The microphone's electrical signal, like the vibration of your speech, is continuous (part **A** in the diagram). The sound card measures the vibration at thousands of points each second. Each measurement is just one moment in time (one dot in **B**), but together they show the shape of the vibration. Each measurement is represented by a number, and the numbers (**C**) are sent to NaturallySpeaking to analyze.

NaturallySpeaking performs many calculations on this stream of numbers as it seeks to determine what you said. The program screens out changes in your voice that aren't useful for recognizing speech. It adjusts the sound signal so that soft words and loud words are treated the same. It also adjusts for the pace of your speech, so that words said rapidly can be recognized by

the same methods as words spoken slowly. The software also filters out static and background noise as best it can.

Spoken words are made up of syllables, syllables are made up of short sounds called phonemes, and phonemes are made up of still smaller “sub-phonemic” parts. When you trained NaturallySpeaking to recognize your voice, you allowed the program to model how you, in particular, say all these phonemes and sub-phonemic parts. NaturallySpeaking analyzes the “cleaned-up” numbers from the sound card by comparing them to the basic components of your voice, seeing what speech components match best. The program uses several techniques, including a mathematical tool called Markov Modeling, to seek these key speech components in the sound of your voice. The software then searches for English words that match closely, using a dictionary of tens of thousands of words stored in the computer’s active memory (RAM).

### **Words That Go Together**

The sound of your voice is only part of the information NaturallySpeaking uses. Just as important is a statistical model of what words tend to go together. This model allows the program to distinguish between words that sound the same or similar, like “which” and “witch,” or “computer is” and “computers.” NaturallySpeaking assumes that the words you’re saying are grammatical—the word “the” will be followed by a noun, and so on. The software doesn’t know actual grammar rules, like following articles by nouns. NaturallySpeaking’s “assumptions” were found inductively by analyzing millions of words of English text. The software knows that after you say “the,” certain other words are more likely to be said next and other words are not likely to be said next.

NaturallySpeaking combines its calculations on the sound of what you said with its estimates of what words tend to go together. It then generates a list of guesses of what you said, in order of certainty. The program types its best guess on the screen. You can see its other guesses as the alternatives in the Correction window.

*Source for information in this section: “When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding,” by Raymond Kurzweil, in Hal’s Legacy, David G. Stork, Editor (Cambridge, MA: The MIT Press, 1997).*

## Why Are NaturallySpeaking's Mistakes So Funny?

---

When NaturallySpeaking makes a mistake, it often types out something that's grammatical. Its guesses are not random, as NaturallySpeaking uses statistical models of what words go together in typical English. The software's mistakes thus have the form of regular writing. While reading random words might be dull and perplexing, reading grammatical sentences engages our human intelligence automatically. NaturallySpeaking's bloopers are similar enough to regular writing to provide a context for understanding them, but different enough from real writing to make us howl. It's the old party game "Mad Libs" updated for the computer age! Some examples are provided here for your reading pleasure.

### Real-Life Bloopers

- ▶ I gave the command, "Bring Up Internet Explorer." I was in a rush to go out and just wanted to check one thing. What Dragon typed was "enough internet."
- ▶ When I try to say "Scratch That" too fast, it types "stress that."  
—Judy L., Sherman Oaks, Calif.
- ▶ When I was writing to one of my penpals I was recovering from a cold. I explained that "I am still sniffing a bit, but at least NaturallySpeaking understands me again." I couldn't have been more wrong...according to the program I was "still sniffing addict." I saw this mistake only *after* I sent the letter.  
—Janneke den Draak, The Netherlands
- ▶ I said, "If you go down to the woods today you're sure of a big surprise." The computer typed, "...you're sure of a big soprano."  
—Derek Fawell, U.K.
- ▶ At the last two Dragon company Christmas parties, a group of us led everyone in singing "The 12 Days of Christmas." We made it a point beforehand to dictate the 12 days of Christmas into different speech recognition

engines. If the phrase came out right (rarely) or was boring, we would say it again and again until we got some funny things to say. When you sing them, the words really do sound like the actual lyrics to the song.

12 Robbers Coming  
11 Diapers Piping  
To Endorse Sleeping (slur the 'to' and 'en-' together)  
Vanity Sensing  
A Tomato Melting  
7 Swanson's Plumbing  
Sixties Delaying  
5 Golden Rings  
4: Birds  
3 French Hands  
2 Turtle Doubts  
and Departure to Repair Tree

—Jeff Foley, Dragon Systems, Newton, Mass.

### **Send Us Yours**

Send us your favorite bloopers! We'll publish the best on our Web site or in a future edition of this book. Bloopers and explanations may be edited for publication. Send bloopers to: Say I Can, 2039 Shattuck Ave. Ste. 500, Berkeley, CA 94704. Or via e-mail to: [editor@SayICan.com](mailto:editor@SayICan.com).